The Distribution of Income Shocks during Crises: An Application of Quantile Analysis to Mexico, 1992–95

William F. Maloney, Wendy V. Cunningham, and Mariano Bosch

Moving beyond the simple comparisons of averages typical of most analyses of household income shocks, this article employs quantile analysis to generate a complete distribution of such shocks by type of household during the 1995 crisis in Mexico. It compares the distributions across normal and crisis periods to see whether observed differences were due to the crisis or are intrinsic to the household types. Alternatively, it asks whether the distribution of shocks during normal periods was a reasonable predictor of vulnerability to income shocks during crises. It finds large differences in the distribution of shocks by household types both before and during the crisis but little change in their relative positions during the crisis. The impact appears to have been spread fairly evenly. Households headed by people with less education (poor), single mothers, or people working in the informal sector do not appear to experience disproportionate income drops either in normal times or during crises.

Mexico’s economic collapse in 1994–95 led to massive declines in household incomes averaging roughly 30 percent. Though the “Tequila crisis” enjoyed a singular celebrity due to Mexico’s proximity to the United States and the comprehensiveness of its prior reforms, neither the large size nor the adverse social consequences of the impacts were unique. Within Latin America, Argentina and Colombia are presently in equally brutal downturns, and in the 1990s

William F. Maloney is lead economist, Office of the Chief Economist for Latin America and the Caribbean at the World Bank; his e-mail address is wmaloney@worldbank.org. Wendy V. Cunningham is senior economist, Human Development Sector, Latin American and Caribbean Region at the World Bank; her e-mail address is wcunningham@worldbank.org. Mariano Bosch Mossi is a junior professional in the Office of the Chief Economist for Latin America and the Caribbean at the World Bank; his e-mail address is mboschmossi@worldbank.org. This article was prepared as a background paper for the regional publication Securing Our Future in a Globalized Economy and was partially funded by the Regional Studies Program of the Office of the Chief Economist for Latin America and the Caribbean at the World Bank. It was also an input into the World Bank report “Income Risk, Household Coping Strategies and Income Security Policy in Mexico.” The authors thank Ishan Ajwad, Omar Arias, Francisco Ferreira, and Norman Hicks for insightful comments. They are also grateful to the Mexican National Institute of Statistics, Geography, and Information (INEGI) for use of their data. INEGI is in no way responsible for any incorrect manipulation of the data or for any erroneous conclusions.

Russia and several countries in Asia experienced notable collapses. Thus, identifying who is most affected in these types of crises is likely to be a central issue in the design of safety nets throughout the world.

The causes of the Mexican crisis have been described at length elsewhere (see Gil-Diaz and Carstens 1996; Edwards 1998; Calvo and Mendoza 1996; Kamin and Rogers 1996; among others). Broadly speaking, liberalization of trade and capital markets in the mid-1980s was followed by an expansion of private borrowing and aggregate demand more generally and by an appreciating exchange rate, all in the context of an underregulated financial sector. Speculative pressure on the peso in late December 1994 led the central bank to allow it to float, and it stabilized at nearly double its previous value against the dollar. Prices rose 35 percent, and output fell 6.2 percent across 1995, with dramatic social consequences. Wages remained fixed in nominal terms, leading to a real decline of 25–35 percent, and unemployment, though low by global standards, almost doubled from 3.9 percent to 7.4 percent (figure 1).

Using a rich panel data set, this article investigates which types of households suffered the largest income shocks during the crisis. It is thus similar in spirit to Gliewwe and Hall’s (1998) study of consumption falls during Peru’s crisis of 1985–90 and is related more generally to the emerging literature on vulnerability.

However, a central theme of this article is that policy analysts need to move beyond estimates of average falls in income or consumption and focus on the entire distribution of positive and negative income shocks. As an extreme example of why this may be important, consider the hypothetical distributions of the income shocks of two groups whose median income change has the same value, $m$ (figure 2). In group A, real income falls more or less evenly across most members, as illustrated by the relatively tight and symmetric distribution of shocks affecting this group. But in group B some heads of households lose their jobs and experience a very large negative shock to household income whereas others experience more moderate negative or even positive shocks, as illustrated by a larger mass of the distribution to the right of the median and a long and thick tail to the left. Looking only at the central tendency of the data therefore obscures important information about who suffered most from the crisis and which groups deserve more attention from a policy perspective.

1. Though there remains ample room for discussion about which particular fundamentals were central, or even whether the crisis was driven by self-fulfilling expectations only loosely linked to fundamentals (Sachs and others 1996), what is important is that commentators do not consider it unique, and so its impacts are of interest for understanding crises more generally. Edwards (1998), for example, sees a strong parallel in the deficient credibility of the reforms in both the Mexican crisis of 1982 and the Chilean crisis of 1998. An alternative view faults excessive borrowing in an environment of unfounded optimism about the future course of the economy. See, for example, Conley and Maloney 1995.
**Figure 1.** Unemployment and Real Wages in Mexico, 1987–2002

*Note:* The real wage is the nominal manufacturing wage deflated by the consumer price index. Unemployment figures are the official numbers covering the entire economy.

*Source:* Mexican Institute of Statistics, Geography, and Information.

**Figure 2.** Use of Quantile Regression as a Tool for Exploring the Distributions of Changes in Household Income

*Source:* Authors’ analysis; see text.
Quantile analysis offers a useful tool for uncovering this hidden information because it permits studying the distributions at several points, not just at the median. The results show significant differences in the distribution of income shocks by household type in Mexico, suggesting that such exploration is indeed valuable. Of perhaps secondary importance, the results also show that median or quantile regression is more robust to the extreme values that commonly emerge in this kind of exercise than traditional ordinary least squares (OLS) estimators and should probably be used routinely in estimating income changes.

The second central theme of the article is that distributions of income shock across households during crises need to be compared against distributions during “normal” times, to see whether the observed differences are related to the crisis or are perhaps intrinsic to particular household types. This question can be turned around to ask whether looking at the distribution of income shocks during normal periods can help identify how vulnerable certain households might be to severe income shocks during a crisis. In the Mexican case, the relative distributions do not change much across periods, with a few important exceptions.

Finally, the analysis shows that several common stylized facts about who experiences the largest negative shocks during crises—particularly about how households headed by disadvantaged single mothers and informal sector workers are affected—are not supported by the data. The article offers some suggestive evidence on why these findings might be reasonable.

I. Quantile Analysis: Studying the Entire Distribution of Income Shocks

Conditional mean regression estimators such as OLS are traditionally used to estimate linear relations among variables. Minimizing the squared sum of errors allows estimating the values of the parameters that predict the mean of the dependent variable, conditional on the chosen set of explanatory variables. However, asymmetries or heteroscedasticity in the distribution of errors may lead to substantially different estimates of the impact of the variables under study at different parts of the conditional distribution. Looking only at the central tendency (the mean, for instance) of the data may thus hide important elements of the story. Furthermore, if there are outliers or if the distribution of the disturbances is nonnormal, mean estimators may be inefficient and biased.

These concerns can be partially addressed by estimating the conditional median regression, in which half the errors lie below and half above the fitted curve. Quantile analysis, introduced in Koenker and Bassett (1978), extends this analysis to estimating curves where approximately $t$ percent of the residuals lie

below the regression line and \((100 - t)\) percent above. Thus, the \(t\)th quantile of
\(Y\) conditional on \(X\) is given by

\[
Q_{Yi}(\tau \mid X_i) = \beta(\tau)X_i,
\]

where \(\beta(\tau)\) is the slope of the quantile line and thus gives the effect of changes in
\(X\) on the \(t\)th conditional quantile of \(Y\). Estimation for different values of \(t\) (from
0 to 1) yields regression lines for various percentiles of the conditional distribution
of \(Y\). Median regression \((\tau = 0.5)\) gives the same results as \(\text{OLS}\) when the
distribution is symmetric.

The analysis presented next suggests that distributions of income shocks
to households show strong evidence of heteroscedasticity and asymmetry and
thus that \(\text{OLS}\) is inappropriate. The following straightforward specification is
designed purely to capture the differential effects of a crisis on the income of
different household types:

\[
\ln Y_{i\tau} = \alpha(\tau) + \sum \delta_j(\tau)D_{ij},
\]

where \(\Delta \ln Y_i\) is the change in the log of per capita labor income of household \(i\), \(D_{ij}\) is
an indicator variable that takes a value of 1 for family \(i\) if it is a member of the
\(j\) different household types at the beginning of the period and a value of 0
otherwise. These groups capture the education level and age of the head of the
household, household composition, and work status of the head. As in standard
\(\text{OLS}\), conditioning on all these categories at once allows disentangling their separate
effects—whether, hypothetically, informally employed household heads might be
particularly hard hit at any quantile relative to the base group (see section II for
details on the base group) captured by the constant, \(\alpha\), or whether it is just that
poorly educated people, who are disproportionately employed in the informal
sector, are hit harder. In the simple case illustrated in figure 2, group A might be
the base group, and the way group B differs from group A would be captured by
including a single dummy variable \(D_{iB}\) to distinguish each household in that group.

Quantile analysis differs from traditional analysis by allowing the entire dis-
tribution of shocks for each category to be parameterized. For group A both the
80th quantile, capturing the less negative and positive tail of the distribution, and
the 20th quantile, capturing the extreme negative shocks, are relatively close to the
median. For group B not only is the variance greater—both the 20th and 80th
quantiles are further from the median than for group A—but the 20th quantile is
displaced far to the left, reflecting that some households experienced extreme
losses such as would arise from loss of employment of the household head.

The values of the coefficients \(\delta_j(\tau)\) in equation 2 at each quantile permit these
different distributions to be sketched out. As in standard \(\text{OLS}\) regressions, at the
median \((\tau = 0.5)\) a negative and significant coefficient on \(D_{iB}\), the dummy variable
representing, for instance, those with incomplete primary education, would imply
that the distribution of shocks to this group is centered to the left of (lower than) that
of base group A in figure 2: “On average” the least educated were hit harder than the
base group. The other two quantiles help flesh out the rest of the distribution. At the 20th quantile a negative and significant coefficient would imply that the value at which the most negative 20 percent of shocks occurred is also shifted down relative to that for the omitted category: The less educated experience lower lows, and hence the tail of distribution B is extended to the left. The coefficient on the 80th quantile tells whether the group experienced higher highs than the base group—whether B has a longer upper tail than A. Together, these three coefficients describe the shape of the distribution of shocks for each group relative to the base group.

Two sets of tests are used to see whether these coefficients are in fact significantly different from each other. First, the coefficients are plotted at each decile along with their standard errors (Koenker and Hallock 2001). Graphically, it is then possible to detect whether the coefficient of one quantile lies outside the confidence interval of another. More formally, an F-test is employed to determine whether the coefficients on the 20th, 50th, and 80th quantiles are statistically different (Gould 1997).

The log difference specification in equation 2 is used because measurement of the magnitudes of change is neutral to the direction of change. The potential downside is that it is only a close approximation to the percentage change for small changes in Y. Because the primary interest is not the total change but the difference from the base category, the log approximation may turn out to be reasonable. However, Kennedy (1981) also shows that for the case of a semilogarithmic estimation like this one, calculating the true percentage change corresponding to the dummy regressors requires transforming the estimated coefficients by

\[
\delta^*_j = \exp(\delta_j - [1/2]V[\delta_j]) - 1,
\]

where \(\delta_j\) is the estimate from equation 2, \(V[\delta_j]\) its estimated variance, and \(\delta^*_j\) the estimated “corrected” value. This turns out to have a significant impact (about 20 percent) on the estimates of the base group changes, which show magnitudes of around 50 percent. As suggested, however, the adjustments to the estimates of the differential effects are generally small. The corrected values are reported here.

Quantile analysis offers two other advantages over traditional techniques. First, by construction, the dependent variable and so the residuals are unlikely to be normally distributed. In fact, were percentage changes to be used, as is

3. Halvorsen and Palmquist (1980) show that in the case of semilogarithmic regressions, when a variable in logs is regressed on a set of dummy variable, the inability to differentiate across the noncontinuous dummy variables means that their coefficients are biased estimators of their impact on the percentage change of the dependent variable. In the case of one dummy variable, \(\ln(Y_2/Y_1) = \alpha + \delta D\) is equivalent to \(Y_2/Y_1 = \exp(\alpha)(1 + \delta^*)^D\), where \(\delta^*\) is the relative effect on the dependent variable of the presence of the factor represented by the dummy variable. Thus, the corrected coefficient that we are interested in is \(\delta^* = \exp(\delta) - 1\). Kennedy (1981) argues that in estimating \(\delta^*\) one must take into account that \(\exp(\delta) = \exp[\delta + 1/2V(\delta)]\), providing the rationale for equation 3.

4. The authors are grateful to Omar Arias for bringing this point to their attention. A previous application of this technique to wage data in Latin America is in Arias (2001).
often done, the distribution would be bounded below at $-1$ and might have very long right tails.

Second, median or quantile regression deals better with extreme values or outliers than do traditional regression techniques. Some very large percentage increases in income are found within the right tails. They might reflect measurement error, but they might also reflect a very low income in the denominator in the first period and a higher income in the second. This might be the case, for example, with self-employed workers, who show very high volatility in their incomes in general. Simple averages will be strongly affected by such values and standard OLS even more so because it minimizes the squared residuals. But median or quantile regressions do not because they are not based on the distance of the residual from the regression line but simply on the number of observations on either side. More generally, this makes median or quantile techniques desirable when outliers or extreme values are present.

As a final methodological point, the very heteroscedasticity that quantile analysis is supposed to reveal also implies that standard errors of Koenker and Basset’s (1978) original formulation are underestimated and thus that the $t$-statistics are overstated (Rogers 1993). As in Gould (1992, 1997) bootstrapping techniques are used to generate the correct standard errors, and Davidson and Mackinnon’s (2000) algorithm is used for determining the correct number of bootstraps.

**What Is the Link to Vulnerability?**

As numerous analysts have noted, there is substantial variance in use of the term vulnerability. In most cases what distinguishes the concept from poverty or deprivation, which are based on the first moments of income or consumption, is its focus on the second moment, the variance, of these measures. It is not just where a household is now but the likelihood that it may find itself in a worse position. This immediately moves the description of the ex ante distribution of shocks to center stage.

Quantile analysis can be an important tool for studying vulnerability to income shocks or to welfare losses where inference based on the assumption of a common distribution of shocks (as opposed to situations such as those described in figure 2) can be misleading. Such concerns are not obviated by the use of limited dependent variable techniques that look only at the probability of movements into poverty. If there is substantial heteroscedasticity in the errors, traditional logit- and probit-based approaches to identifying vulnerability to becoming poor will generate inconsistent estimates (Yatchew and Griliches 1984). This offers a rationale for analysis in a continuous context (Glewwe and Hall 1998; Cunningham and Maloney 2000).

Repeated panels can be used to estimate the distribution of shocks during normal times for each household type. However, it is not obvious that this

---

information is useful for identifying who is likely to experience large income shocks during crises, when almost by the definition of crisis the underlying data generation process may have changed. This is explicitly tested in the Mexican crisis in the next section. In the event that the distributions change and prior information is therefore not useful, Glewwe and Hall (1998) and this article can then be seen as offering individual draws from the crisis data generation process. They can tell us who experienced a given shock in a particular event but not who is most likely to experience such a shock in repeated events, which is of greater interest for policy design. Identifying who is most vulnerable to income or consumption shocks during crises would require repeated observations of crises. This, in turn, is likely to be complicated by the fact that different precipitating causes and contexts (for instance, how labor markets adjust) may imply very different patterns of shocks to household income.

II. Data

The analysis studies changes in per capita income of families across the period 1994–96 using the Mexican National Urban Employment Survey (ENEU). The ENEU has conducted extensive quarterly household interviews in the 16 major metropolitan areas from 1992 to the present, which includes the period of the Tequila crisis. The sample is selected to be geographically and socioeconomically representative. The questionnaire is extensive in its coverage of issues traditionally found in such employment surveys, including participation in the labor market, wages, and hours worked. Additionally, a household identification variable permits construction of household incomes.

The ENEU is structured as a rotating panel. Each quarterly sample includes five cohorts, each in a different stage of the interview cycle: one-fifth of the sample in its first interview, one-fifth in its last (fifth) interview, and three-fifths in intermediate stages. To construct the panels and ensure proper identification, individuals were linked by position in an identified household, level of education, age, and gender.6 Household incomes were then constructed by aggregating across the reported household members.

A panel was constructed with 9,877 households beginning in 1994:3 to 1995:3, covering the onset of the crisis and period of largest wage losses around the Tequila crisis. Dummy variables were included for the education level of the head (primary incomplete, primary complete, secondary incomplete, secondary complete); age (under 25 or over 45); more than the mean number of children (1.3) in the household; and household structure (single mothers with children, single women without children, and single men without children). Three dummy variables were also

6. Using just the first variables to concatenate and following changes in sex across the panel led to mismatching (or misreporting) of less than 0.5 percent. Initial and final values have under 10 percent zeros, approximated by adding a 1 prior to logging.
included for the sector in which the household head works (informal self-employed, informal salaried, and a residual category for being out of the labor force, unemployed, or otherwise not earning an income). The constant, \( a \), captures the base group of households headed by married, middle-age, college-educated males working in the formal sector, with less than the mean number of children.

The quantile analysis provides a very complete description of the distribution of the movement of household incomes. Worth highlighting is the relatively fluid movement between income quintiles. For example, only about 30 percent of households were in the same income quintile after five quarters, and a similar share jumped from the bottom to the top two quintiles. There are several possible reasons. First, roughly 40 percent of household heads are self-employed, and as the next sections show, entrepreneurs experience much more volatility in income flows than do workers in the salaried formal sector, reflecting varying business conditions or personal preferences. This higher volatility is also a characteristic of cross-sectional data and thus is not related to the process of linking the panels. Second, there is a very high degree of mobility in the labor force, and frequent movements between formal and informal sector jobs imply large changes in reported income. Third, there may be measurement error arising from the survey process, imperfect recall by the interviewed household member, or noise in the tabulation process. It seems reasonable to assume that this error is not correlated with the household categories and so would not affect the description of each category’s distribution relative to that of the base.

III. RESULTS

The OLS estimates of equation 2 indicate that the income of the base group, captured by the constant, fell about 52 percent over the one-year period (table 1, column a). To find the impact on other groups, and to see whether it differs statistically from the impact on the base group, the value of the dummy variable is added to the constant. Those with an incomplete primary education experienced income falls 36 percentage points less than those of the base group, or under a third of the base group. The income of informal self-employed household heads fell 5 percentage points more than that of those of the base group, or under a third of the base group. The income of informal self-employed household heads fell 5 percentage points more than that of the base group.

7. The term informal is used here to refer to workers unprotected by labor laws. It includes owners of firms with fewer than 16 employees who do not receive social security or medical benefits (fewer than 1 percent have more than five employees) and employees in these small firms, identified as informal salaried workers.

8. Single men with children made up less than 0.7 percent of the sample; they are included in the base group.

9. Maloney (1999) finds that voluntary movements into self-employment from formal salaried work lead to a 30 percent average jump in reported earnings. Some 70 percent of those who moved report moving voluntarily. In deciding what sector to work in, individuals compare welfare, not wages, so a move from a sector with benefits and job security to a microenterprise where firm mortality rates are very high dictates a substantial compensating wage and risk premium. This could be what is picked up in the high apparent mobility among income levels.
of the base group and the income of nonremunerated household heads fell 20 percentage points more. These results are illustrated in figure 3, which charts the OLS coefficient (solid lines) and confidence interval (broken horizontal lines), as well as the estimated coefficients at each quantile from the 10th to 90th (solid curves) along with its confidence interval (broken curves).

The median regression (the 50th quantile) shows a different picture, however (see table 1, column d). The base group experienced a 29 percent drop in income, roughly half the estimate from the OLS regression. Examination of the data reveals that a few outliers are largely responsible for this difference.

Table 1. Correlates of Log Changes in Household Income during the Mexican Crisis: Quantile Analysis, 1994:3–1995:3

<table>
<thead>
<tr>
<th>OLS Regression</th>
<th>Quantile Regressions</th>
<th>F-testb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Adjusted (±3 SD)a</td>
<td>20%</td>
<td>50%</td>
</tr>
<tr>
<td>Primary incomplete</td>
<td>0.363***</td>
<td>0.031</td>
</tr>
<tr>
<td>Primary</td>
<td>0.306***</td>
<td>0.040</td>
</tr>
<tr>
<td>Secondary incomplete</td>
<td>0.185***</td>
<td>-0.05</td>
</tr>
<tr>
<td>Secondary</td>
<td>0.192***</td>
<td>0.016</td>
</tr>
<tr>
<td>Young</td>
<td>-0.014</td>
<td>-0.068</td>
</tr>
<tr>
<td>Old</td>
<td>-0.148</td>
<td>-0.068***</td>
</tr>
<tr>
<td>More than 1.3 children</td>
<td>0.037</td>
<td>0.005</td>
</tr>
<tr>
<td>Single mothers</td>
<td>0.024</td>
<td>0.034</td>
</tr>
<tr>
<td>Single women</td>
<td>0.046</td>
<td>0.009</td>
</tr>
<tr>
<td>Single men</td>
<td>-0.014</td>
<td>-0.072</td>
</tr>
<tr>
<td>Informal self-employed</td>
<td>-0.053*</td>
<td>-0.046</td>
</tr>
<tr>
<td>Informal salaried</td>
<td>0.030</td>
<td>-0.006</td>
</tr>
<tr>
<td>No remuneration</td>
<td>-0.197***</td>
<td>-0.194***</td>
</tr>
<tr>
<td>Base group</td>
<td>-0.517***</td>
<td>-0.360***</td>
</tr>
<tr>
<td>Number of observations</td>
<td>9,877</td>
<td>9,636</td>
</tr>
</tbody>
</table>

*aSignificant at the 10% percent level.
**Significant at the 5% percent level.
***Significant at the 1% percent level.

Note: The dependent variable is the change in log per capita income of the household between 1994:3 and 1995:3 regressed on characteristics of Mexican households and their heads. Standard errors were calculated using bootstrapping techniques from Gould (1992, 1997). Number of bootstraps were obtained by running the algorithm proposed by Davidson and Mackinnon (2000). Coefficients were corrected following Kennedy (1981).

Observations with income growth outside ±3 standard deviations from the mean have been dropped.

The null hypothesis is equality of the coefficients in the three quantile regressions estimated.

Source: Authors’ analysis based on data from one quarter panel of the Mexican National Urban Employment Survey.
When observations beyond plus or minus three standard deviations are dropped, the OLS estimate approaches the median estimates, at 36 percent (column b). Median regression is preferred to such trimming, however, because the trimming cutoff is arbitrary. For example, 74 of 241 observations that are classified as outliers are households headed by self-employed workers whose incomes show high variability even under normal circumstances and hence may be legitimate outliers. An additional 34 observations are in the nonremunerated category, for which a change to even small earnings could increase household income by a very large percentage. Because the median estimator is less sensitive to such extreme values, it avoids the arbitrary discarding of outliers.

Among household categories with significant coefficients in the median regression, households headed by workers with incomplete primary education saw their incomes fall by 4.5 percentage points less than the base group and those with completed primary school by 6.5 percentage points less. Households headed by older workers experienced drops 4.5 percentage points greater than did the base group, and the self-employed 7.1 percentage points greater. That these coefficients are significant is reflected in figure 3 by the fact that the confidence interval at the 50th quantile does not span the zero line.

What also becomes apparent is the value of looking beyond the median to other quantiles. If in figure 3 the parameter value at one quantile lies outside the confidence interval of another, there are statistically significant differences in their values. Thus the impact of the explanatory variable differs across quantiles. As an example, for households with heads with incomplete or complete primary education, the 20th quantile estimates lie above and outside the confidence interval of the 50th, showing that they are significantly greater in value. That significant differences occur is confirmed by the $F$-test (see table 1), which shows that for all education categories, the old, the self-employed, and those with no remuneration, the hypothesis can be rejected that the coefficients at the 20th, 50th, and 80th quantiles are statistically the same.

More careful examination at the 20th quantile—the value below which the most extreme shocks are found—shows that households with heads who are less educated were more cushioned than the base group of households whose heads were college educated, who show an income drop of 66 percent. For households whose heads had a primary or incomplete primary education, the 20th quantile is only around $-15$ percent. For complete and incomplete secondary education, it is roughly $-40$ percent. The negative coefficient suggests that shocks to income were harsher for the old and self-employed by 23 percentage points and by even more for those whose heads are without remuneration.\(^{10}\)

\(^{10}\) Note that the overall effect for this group is a more than 100 percent drop in income, which seems implausible. This is due to the bad approximation properties of the log transformation before great percentage changes (even after using Kennedy's correction), such as when going from some income to no income at all.
The 80th quantile, which completes the distribution, shows few differences from the pattern at the median. None of the education or household composition dummy variables is significant. For all three quantiles the distribution of the poorly educated relative to the base group of college educated does not correspond to what is sketched in figure 2. The distribution overall is shifted to the right at the median and has shorter, lower tails, but it appears to be the same as the base case above the median—it is simply compressed below it. Poorly educated workers, again conditional on all other characteristics, experienced equal or more moderate income shocks than the base group at every point of the distribution.

The most strikingly distinct overall distribution is now that of families with self-employed heads whose coefficient at the 80th quantile is 12.5 percentage points.
Note: These figures represent the coefficients and the 95 percent confidence intervals for a number of quantile regressions (10th–90th) for each of the variables in the estimation. The OLS coefficient and confidence intervals are included for comparison. Obviously, the OLS regression returns a constant coefficient over the different quantile regressions. A zero line in the y-axis is also reported to evaluate statistical significance.

Source: Authors’ analysis based on data from one quarter panel of the Mexican Urban Employment Survey.
higher than that of the base group, suggesting that they not only experienced a
worse median shock, but that their variance is much greater, because they demon-
strate higher highs (at the 80th quantile) and lower lows (at the 20th quantile).
Somewhat surprisingly, the distribution of households headed by the salaried
informal sector workers shows no difference from the base group of formal salaried
workers except that their highs are 7.7 percentage points higher. Similarly, families
with households headed by workers with no remuneration show a coefficient at the
80th quantile that is 11 percentage points higher than that of the base group. The
household composition variables are not significant at any quantile. The distribu-
tions of shocks for families headed by single women, single men, or single mothers
are indistinguishable from those for households headed by married men.

IV. Are “Normal” Periods Good Predictors of Crises?

The logical question to ask is whether these differential patterns of shocks are due to
the crisis or whether they exist in normal times as well. Or, in the context of asking
about the vulnerability of different groups to income shocks, can the distribution
estimated during normal periods be taken as a predictor of what distributions will
look like during crises? The same regressions are run again, this time using a year of
complete panels to form a sample ending right before the crisis (table 2). This is
constructed by combining four additional panels: beginning in 1992:2, 1993:1,
1993:2, and 1993:3, the last of which ends in 1994:4, immediately before the
economic collapse that began after the devaluation in late December 1994.

The results show that the distinct distributions exist in normal periods as well
(first three columns of table 2). Household heads with primary, incomplete
primary, and incomplete secondary education show values at the 20th quantile
shifted to the right relative to values for the base group while the 80th quantile
is shifted to the left—they have lower highs and higher lows although the
median change is statistically the same. Put differently, households with less
educated heads generally have lower variance in their household incomes. The
reasons are not clear. On the upside, perhaps the possibilities for income growth
are lower for the less well-educated. On the downside it is possible that
fewer opportunities to smooth income through credit markets or savings dictate
that households take measures to reduce downside volatility—for example, by
putting additional workers in the labor market if the household head’s job is
lost.

The reverse occurs for self-employed and salaried informal sector workers and
for households whose head is not remunerated. In all three cases the 80th quantile
is shifted to the right and the 20th to the left, suggesting a greater variance in
earnings. The self-employed also have a median income change below that of the
base group. Taken together, this seems consistent with the standard dynamics of a
small firm sector: In any period some firms do exceptionally well, and some less
well than their formal salaried counterparts, and as is the case in many countries,
firm mortality rates are very high. This could be argued as reflecting greater precariousness, but because almost 70 percent of workers entering the self-employed sector from formal salaried work report doing so voluntarily, this greater volatility is consistent with welfare improvement (Maloney 1999).

Table 2. Comparison of Shock Distributions before and during Crisis, Quantile Analysis

<table>
<thead>
<tr>
<th></th>
<th>Presample Period</th>
<th></th>
<th></th>
<th>Crisis Period</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20%</td>
<td>50%</td>
<td>80%</td>
<td>20%</td>
<td>50%</td>
<td>80%</td>
</tr>
<tr>
<td>Primary complete</td>
<td>0.232***</td>
<td>0.009</td>
<td>-0.065***</td>
<td>0.232***</td>
<td>0.036</td>
<td>0.073***</td>
</tr>
<tr>
<td>Primary</td>
<td>0.183***</td>
<td>0.005</td>
<td>-0.055***</td>
<td>0.224***</td>
<td>0.060***</td>
<td>0.066**</td>
</tr>
<tr>
<td>Secondary complete</td>
<td>0.108***</td>
<td>0.002</td>
<td>-0.033**</td>
<td>0.128*</td>
<td>0.012</td>
<td>0.039</td>
</tr>
<tr>
<td>Secondary incomplete</td>
<td>0.079**</td>
<td>0.008</td>
<td>-0.016</td>
<td>0.166*</td>
<td>0.023</td>
<td>0.028</td>
</tr>
<tr>
<td>Young</td>
<td>0.041</td>
<td>0.008</td>
<td>-0.029</td>
<td>-0.102</td>
<td>-0.009</td>
<td>0.019</td>
</tr>
<tr>
<td>Old</td>
<td>-0.192***</td>
<td>-0.047***</td>
<td>0.014</td>
<td>-0.046</td>
<td>0.001</td>
<td>-0.039</td>
</tr>
<tr>
<td>More than 1.3 children</td>
<td>0.028</td>
<td>0.017</td>
<td>0.004</td>
<td>0.019</td>
<td>-0.005</td>
<td>0.012</td>
</tr>
<tr>
<td>Single mothers</td>
<td>-0.047</td>
<td>0.005</td>
<td>0.008</td>
<td>-0.001</td>
<td>0.030</td>
<td>0.031</td>
</tr>
<tr>
<td>Single women</td>
<td>-0.163***</td>
<td>-0.055***</td>
<td>-0.079***</td>
<td>0.219***</td>
<td>0.061*</td>
<td>0.116***</td>
</tr>
<tr>
<td>Single men</td>
<td>-0.103*</td>
<td>-0.030</td>
<td>-0.001</td>
<td>0.031</td>
<td>0.026</td>
<td>0.014</td>
</tr>
<tr>
<td>Informal complete</td>
<td>-0.263***</td>
<td>-0.069***</td>
<td>0.105***</td>
<td>0.049</td>
<td>-0.005</td>
<td>0.018</td>
</tr>
<tr>
<td>Informal self-employed</td>
<td>-0.093***</td>
<td>0.020</td>
<td>0.077***</td>
<td>0.071</td>
<td>-0.008</td>
<td>0.000</td>
</tr>
<tr>
<td>Informal salaried</td>
<td>-0.157***</td>
<td>0.059***</td>
<td>0.242***</td>
<td>-0.447***</td>
<td>-0.124***</td>
<td>-0.106***</td>
</tr>
<tr>
<td>No remuneration</td>
<td>-0.436***</td>
<td>-0.018*</td>
<td>0.529***</td>
<td>-0.406***</td>
<td>-0.280***</td>
<td>-0.308***</td>
</tr>
<tr>
<td>Base group</td>
<td>41,676</td>
<td>41,676</td>
<td>41,676</td>
<td>41,676</td>
<td>41,676</td>
<td>41,676</td>
</tr>
</tbody>
</table>

*Significant at the 10% percent level.
**Significant at the 5% percent level.
***Significant at the 1% percent level.

Note: The table shows the coefficients of a regression in which five periods (1992:4–1993:4, 1993:1994:1, 1993:2–1994:2, 1993:3–1994:3, and 1994:3–1995:3) of household per capita income growth have been regressed on characteristics of Mexican households and their heads. The regression includes dummy variable interacted with one of each variable for the crisis period that goes from 1994:3 to 1995:3. The second column for each period retrieves the coefficients of the interaction of each variable with the dummy variable for the crisis period. Standard errors were calculated using bootstrapping techniques from Gould (1992, 1997). Number of bootstraps were obtained by running the algorithm proposed by Davidson and Mackinnon (2000). Coefficients were corrected following Kennedy (1981).

Source: Authors’ analysis based on five quarter panels of the Mexican National Urban Employment Survey.

11. See Levenson and Maloney (1998). Fajnzylber and others (2003) find that the rate of micro-enterprise owners returning to formal employment, one measure of enterprise mortality rates, is roughly equivalent in Mexico to that in the United States.
The same lower than base group growth in median incomes occurs for households headed by older workers. This may reflect a tendency toward declining incomes with retirement. The reverse logic may hold for families whose head shows no earnings at the beginning of the period. On average, many of those unemployed will find jobs, so the tendency of this group to increase its incomes above the median for the base case is as expected.

Households headed by single women have their entire distribution shifted to the left, doing worse at every quantile. The pattern is similar for single men, although the coefficients are insignificant for the upper two quantiles. This may suggest that jobs held by young people generally have fewer possibilities for large gains, greater possibilities for layoffs, and mediocre performance at the median.

Households headed by single mothers, on the other hand, appear to have the same distribution of shocks as does the base group. Though perhaps surprising, this finding is consistent with anthropological studies of Mexican families. Selby and others (1990) find that matrifocal families have higher per capita incomes, proportionally more family members in the workforce, and lower dependency ratios and that these families generally do as well as nonmatrifocal households. In fact, Chant (1985, p. 650) finds that “despite major structural constraints of the economic and social potential of matrifocal families, single parent units frequently fare better than male headed households.”

As a final observation, larger households tend to have higher income gains than smaller families. This may reflect greater effort by the household head or labor force entry by additional family members who contribute to the family pot.

These findings require revisiting the conclusions of section III. The values and significance levels of interactive terms corresponding to the crisis period for the combined precrisis and crisis periods are shown in the last three columns of table 2. These coefficients can be interpreted as showing how group X did relative to the base group in normal periods compared with how group X did relative to the base group during the crisis. The simplest example is that the

12. “Matrifocal households are not worse off because of the absence of a man, despite discrimination against women in the work force and the difficulties that women have getting well paid employment. Although median household income is 14% lower than non-matrifocal households, since they average one less member in the household, their per capita incomes are 8.2% higher. They put almost as many members into the paid work force (1.38 vs. 1.4) and the ratio of dependent to members in the work force is lower than the non-matrifocal… The genealogical complexity of matrifocal households is quite outstanding, with more grandchildren, more siblings, and more outsider as members. Though they are smaller, they are not abandoned and alone and preyed upon by society. Most of them are viable, operating units, doing as well as non-matrifocal households in the very tough world of the ‘popular classes’ of urban Mexico… We had thought that they would be highly vulnerable and living on the margin of existence. We had been persuaded of this notion by interviews with people from nuclear or extended families, and case studies, carried out mostly in Oaxaca, of families in tragic circumstances, recently abandoned by a drunken and abusive male, cast upon their own resources, or eking out a pitiable existence taking in washing and making tortillas to sell to other households. To our great relief, it seems that such circumstances are relatively transitory, particularly if people are long-enough established to have kin-folk upon whom they can rely” (Selby and others 1990, p. 95).
coefficient on the incomplete primary education group in the crisis period of table 1 is the sum of the coefficient in the normal period in table 2 and the dummy interactive term in the second panel.\textsuperscript{13} Not surprisingly, the results for the crisis period in table 2 reveal a shifting downward of the distribution of the base group, with median and 80th quantile following roughly equivalently, whereas the lower tail captured by the 20th quantile elongates significantly, perhaps capturing a rise in job loss.

The interactive terms on the non–base group dummy variables can be seen as adjusting their distribution relative to the adjustment in the distribution of the base group. If there are no significant interactive terms on group \( X \), that means that the distribution of the group changed in the same way that the base group’s distribution changed.

A revealing example is that none of the interactive terms is significant at any quantile for the self-employed. This means that the findings in section III of higher highs, lower medians, and much lower lows for this group were not a function of the crisis because the same results are found in normal times. It cannot be rejected that the distribution of this group relative to the base group remained statistically equivalent. In contrast to the previous findings, they cannot be said to have suffered unusually during the crisis. The same can also be said for informal salaried workers.

Similarly, the distributions for the old and young, those with above-average-size families, single mothers, and single men track the base group, suggesting that these groups also do not appear to have suffered any more or less in either median or variance than did the base group. The performance of single mothers is again somewhat surprising, yet consistent with Glewwe and Hall’s (1998) findings for Peru.

The same cannot be said of families with heads who earned no income. Their distribution shifted sharply left at every quantile relative to the movement of the base group. Although both median and 80th quantile were previously higher than for the base group, perhaps reflecting that the head of the household often got a job across the sample period, this effect is reversed at the median and sharply attenuated at the 80th quantile during the crisis, perhaps reflecting the increased difficulty of getting a job.

Most striking is that the less educated the household head, the less likely the distribution is to follow the college educated base group to the left during the crisis. For heads with a primary or an incomplete primary education, the value at the 20th quantile falls only 17 percent, compared with 40 percent for the base group, and at the 80th quantile income falls only 23 percent compared with 30 percent for the base group. The less well-educated appear to do better during the crisis than in normal periods relative to the base group, experiencing lower lows, higher highs, and, for those who completed primary school, a higher median. The premium at

\textsuperscript{13} Because of the corrections discussed earlier, the summation is not always exact and often diverges significantly at the extreme quantiles.
the median found for this group in table 1 is, in fact, a feature of the crisis. This may be due to a greater propensity for poorer families to put extra workers in the labor force during crises. The same effect appears, although diminished in magnitude and statistical significance, for workers with a secondary education.

V. Conclusions

Quantile analysis was used to identify which households suffered the largest income falls during the 1995 crisis in Mexico. The analysis provides estimates of “average” shocks, which are more robust to outliers and the nonnormality of the distribution and more fully describes the heteroscedasticity and asymmetries in the distribution of shocks, which appear to be important in understanding who experiences the most extreme falls in income. Because the underlying distribution of shocks is central to the concept of vulnerability more generally, these tools are potentially very useful.

The results suggest several stylized facts. First, even during normal times different groups have very different distributions of shocks. As an example, households headed by informal self-employed workers have great variance in incomes, consistent with a sector made up of small businesses, which experience great volatility. Yet most workers voluntarily enter the self-employed small business sector, suggesting that taking on this high variance is consistent with welfare improvement. Thus, in attempting to measure job quality, the second moment of incomes needs to be included as only one of numerous job characteristics considered in workers’ decisions on sectoral choice.

Second, these normal-period distributions of income need to be taken into account when evaluating the impact of a crisis. Once they are controlled for, households that enter the crisis with a head who is unemployed do substantially worse in the distribution of outcomes than during normal periods and compared with the change in the distribution of the base group. However, virtually every other group fared as well or better than they did in noncrisis times relative to how the base group did. The least educated appear to have some slight gains at the median and very substantial gains at the lower tails of the distribution—their negative shocks are smaller than those of the base group. Neither informal workers nor households headed by single mothers experienced especially severe shocks at any point of the distribution, even though for self-employed workers that appears to be the case when only the crisis period is examined.

These differences in the relative shapes of the noncrisis distribution from the crisis distribution suggest that it is problematic to use normal-period distributions to forecast who would be most vulnerable to large income shocks during crises. Furthermore, different precipitating causes and contexts may imply very different patterns of shocks during crises and thus even greater differences between the two. For instance, during the Mexican crisis labor markets adjusted primarily through wages, not quantities. Most wages were held fixed in nominal terms while inflation eroded their real value, and unemployment did not become
especially high by regional standards. This is consistent with the finding here that the costs of the crisis were spread relatively evenly across household types—thus the broadly similar movements in the distributions across groups from precrisis to crisis periods. This may not be the case during crises in which labor markets adjust through quantities and unemployment is not distributed evenly across household types.

Finally, although it is worthwhile knowing which types of household experience the largest shocks to income during a crisis, it is not easy to move from there to statements about welfare. If the poor have fewer savings or if they cannot borrow, a given income loss may lead to larger consumption falls than for the rich. The Mexican income shocks from the crisis were permanent in the sense that incomes did not begin to recover for at least three years, so such income-smoothing strategies would be ineffective in the long run: Consumption would follow income reasonably closely. Even if income shocks led to identical consumption losses, poor families might be less able to tolerate these than the better off, particularly if they are forced below the poverty line. Ideally, there would be some mapping income shocks to welfare, either in a continuous context or as is implicit in poverty line analysis.

REFERENCES


