Assessing Trends in Informality in Argentina: a Cohorts Panel VAR Approach

Omar Arias
The World Bank

Walter Sosa Escudero*
Universidad de San Andrés and CEDLAS/UNLP

This version: May 18, 2007

Abstract: This paper studies the interrelation between labor informality, relative informal/formal wages and unemployment in Urban Argentina during the last three decades using Panel VAR methods. The very short time span of available household level panel data is overcome by constructing synthetic cohorts of archetypical individuals that can be followed over a longer period, and that also help mitigate measurement errors in individual earnings. The results are consistent with exclusion or segmentation as the main driver of informal salaried employment, while independent employment is largely driven by voluntary choice of workers (that is relatively persistent) although with a minor subgroup of workers that maybe excluded from more desirable formal salaried jobs (which moves counter to the economy cycle).

JEL Classification: C12, C23, C52
Keywords: informality, Argentina, synthetic cohorts, panel vector autorregression.

*Corresponding author: Walter Sosa-Escudero, Department of Economics, Universidad de San Andres, Vito Dumas 284, B1644BID Victoria, Argentina, Ph/Fax: (54-11) 4725-7000, e-mail: wsosa@udesa.edu.ar. Jorge Balat and Javier Alejo provided excellent research assistance. We thank Inessa Love for kindly making available her PVAR routines and Leonardo Gasparini for allowing access to many software and data resources from the Centro de Estudios Distributivos Laborales y Sociales (CEDLAS). The usual disclaimer applies.
1 Introduction

Labor informality in developing countries is perceived as an obstacle to greater productivity, higher per capita incomes, and a constraint to the State’s capacity to collect taxes and provide key public goods, trapping workers in a vicious circle of low productivity and precarious employment. As in many other Latin American Countries, informality matters in Argentina because of its levels and its evolution. The country has seen the largest increase in informal employment in the region over the last decades. The fraction of all workers who are informal employees (lacking social security coverage in their current job) has grown significantly in the last two decades, by 17.4 point for the Buenos Aires Metropolitan Area during 1980-2003 and by 8.6 points in urban areas from 1992-2003. Meanwhile, independent employment has remained relatively more stable. At the end of 2005, 35% of the workers were informal employees and 23% were independent workers. The increase in informality has been largely concentrated among low-income households and medium size firms, deepening inequality, and informal work is now the dominant form of employment among workers in poor households. A deep understanding of this evolution is essential to implement policy measures aiming at improving social and economic welfare, and also because it sheds light on how labor markets react in economies subject to sudden changes in its underlying rules and conditions, including expansions and crisis, and profound changes in its structure.

There is a vivid debate on whether informal workers are a particular subset of the disadvantaged sectors of the society or whether, instead, these workers are there largely voluntarily. The first perspective corresponds to the ‘exclusion’ or segmentation view that sees the informal sector as a refuge for workers who are unable to find desired formal salaried jobs, that is, the disguised unemployed. In this view, workers are pushed towards informality. The second perspective corresponds to the ‘integrated’ or voluntary view that sees informality as a dynamic unregulated sector which generate jobs that are not necessarily inferior. In this lense, workers are pulled to the informal sector attractive by competitive
earnings prospects, desires of independence, flexibility, and the low opportunity cost of formal jobs given low formal sector productivity and access to other means of social protection (private safety nets or universal services that do not require contributions).

The possible causes for the rise in informality in Argentina are many and still poorly understood. Many observers point to macroeconomic factors (changes in exchange rate and price regimes), structural reforms (trade opening, privatization, etc.), changes in labor regulations (i.e., flexible employment contracts), or the decline in union power. Existing studies do not find robust evidence for any of these to be the silver bullet, and rather suggest a possible modest contribution of each (World Bank 2007). Noteworthy, Bour and Susmel (2000) note that informal salaries declined relative to formal salaries during the 1990’s. They point out that the Convertibility period was characterized by low inflation and even deflation, and argue that this made formal sector salaries stickier. In other words, during a period of recession and unemployment, formal sector salaries cannot easily decline in nominal terms and consequently employment shifted to the informal sector, where nominal salaries can adjust downward more easily. This suggests a trade-off between unemployment and informal employment for the economy. However, informality rose at the same time that unemployment was rising and even when the economy experienced significant growth in the early 1990s.

As Fiess, Fugazza and Maloney (2002, 2006) point out, these dynamic interactions can be learned by studying the time series properties of informality rates and wage differentials, together with other macroeconomic variables. However, at the micro (household or individual) level there is a natural limitation given by the absence of panel data that covers long periods. Even in the case of Argentina with a long tradition carrying out a regular household survey, the Encuesta Permanente de Hogares (EPH), only allows to construct rotating panel where households can be followed in four time points along two years. This is the approach adopted in Arango and Maloney (2000) and Pages et al (2006) who exploit the short panel structure of the EPH to study labor and earnings
transitions. A different strategy is the one adopted by Fiess et al. (2002, 2006) who work at the national level with a panel of four countries (Argentina, Brazil, Colombia and Mexico) to study the comovements between informality rates, relative incomes and exchange rates using standard time series techniques.

In this study we propose an intermediate empirical strategy relying on synthetic cohorts of individuals. We construct a pseudo-panel dataset which preserves many of the microeconomic structure of informality (age or educational profiles) while allowing to identify temporal and cohorts effects, including structural breaks, due to the longer-time span of the series which allows the use of time series methods. We will study the comovements of informality rates for groups of individuals defined by age cohorts using a panel VAR (PVAR) approach. As it is well known from the applied macro literature, vector autorregressions provide a flexible statistical framework to study the dynamic interaction of variables when there is no clear theory regarding the causal link among them. In light of the previous discussion, this is clearly the case for relative wages, informality rates, and unemployment where interactions and feedback effects occur without a clear causality direction. This empirical strategy allows us to assess the importance of general cyclical movements, the interactions between unemployment and informality and the responses to exogenous changes in each of them as well as exploring the relevance of structural changes.

Cohorts methods have proved a valuable source of information in dynamic contexts since they provide a good balance between the abundant individual information but short spanned temporal dimension of permanent household surveys, and the longer span of time-series of highly aggregated country level studies. By exploiting the pseudo-panel structure we can mitigate composition biases that can arise in studies using aggregate variables (such as in Fiess, Fugazza and Maloney) due, for example, to changes in the demographic and skills composition of the labor force. This also helps mitigate measurement errors in earnings who can obscure relations between the permanent and true transitory components of interest (Antman and Mackenzie 2005).

The paper is organized as follows. The next section describes the cohorts
approach and section 3 the PVAR strategy. Section 4 describes the data and the construction of synthetic cohorts. Section 5 presents the empirical results. Section 6 summarizes the findings and their implications.

2 The cohorts approach to labor informality dynamics

Many issues related to individual or household dynamics seem to rely crucially on the availability of panel data, and labor informality dynamics is a clear example of this. Understanding how the decision or process that leads to informal employment interacts with market wages and unemployment requires following the same individuals for a relatively long period where these interactions can be learned and measured.

Empirical analysis faces the serious restriction imposed by the fact that in most developing countries household surveys do not have a panel structure but, instead, are collected as as sequence of cross-sections. An intermediate situation arises in cases like Argentina’s Encuesta Permanente de Hogares (EPH, Permanent Household Survey) which has a rotating panel structure, where households are interviewed four times in two years, with one fourth of the families being replaced each period. This provides a valuable source of panel information which has been exploited in many studies, but proves insufficient for our purposes, mostly due to its very limited time span, which covers only two years. As it is well established in the literature, dynamic panel methods are severely affected by a short-time span, that cannot be compensated by the large household or individual dimension. Also, temporal effects like changes in macroeconomic factors or structural breaks, require a rather long temporal window to be properly isolated. In addition, even when panel information from household surveys can be recovered, the fact that these surveys are not specifically designed to provide panel information leaves ample room for all the concerns attached to panels, like non-random attrition, measurement error, as has been clearly stressed by Ashenfelter, Deaton and Solon (1986).
Meanwhile, the study of dynamic relations based on country level aggregates solves the short temporal dimension only partially since in our case the temporal evolution of the EPH can be reliably and continuously constructed only back to the early 1980’s, still a short time span for dynamic analysis. This problem is compounded by the fact that the aggregation process leaves very few degrees of freedom by concentrating the relevant information into no more than 20 yearly observations. Besides, the aggregation process naturally eliminates all the cross-sectional sampling variation by averaging important heterogeneities at the household level, and can give rise to biases if the composition of the populations underlying the national averages change significantly.

The cohorts approach offers a reasonable intermediate strategy which exploits both the temporal and cross-sectional variability. A cohort is simply a group of individuals who experienced a particular event at a point in time.\(^1\) The most commonly used ‘event’ is year of birth, in which case the unit of analysis is no longer individuals but rather ‘individuals born at a specific year’. The obvious advantage is that, unlike individuals or households, these ‘archetypical’ individuals can be followed over time using the age information available in household surveys.

In essence, this strategy relies on a ‘pseudo panel’ where each observation corresponds to a specific cohort in a given year. For example, in our case we will construct informality rates at a specific year for a given cohort, say, the informality rate in 1998 of individuals who were born in 1965. This is done for the period covering 1985-2003. The cross-sectional variation is provided by the coexistence of different cohorts at each period.

The intermediate aggregation process implicit in the construction of cohort observations may help overcome some potentially serious problems present at the individual level, in particular, measurement errors, as clearly stressed by McKenzie and Antman (2005). This is of special concern for measures of individual earnings which are known to be subject to non-trivial measurement errors. The correlation of these measurement errors over time can cloud the

\(^1\) Glenn (2005) and Deaton (1997) are good introductions to the subject.
analysis of the dynamics of earnings. The same may apply to other employment variables, for example informality, where short run random episodes may induce spurious mobility in the informality status of individuals.

The construction of average income or informality rate of a cohort naturally average out these measurement errors, as well as unobserved heterogeneities at the cohort level as the number of individuals in each cohort is large. These ‘pseudo-individual’ heterogeneities at the cohort level are the main source of inconsistencies and biases in dynamic panel data models, and the main reason to resort to instrumental variables strategies, in the spirit of the well known Arellano-Bond (1991) approach. See Mackenzie (2003) and Verbeek and Vella (2005) for this issue, we will refer back to this topic later when we describe the panel VAR approach. To our knowledge, ours is the first application of dynamic cohort methods to the analysis of informal employment, unemployment and earnings dynamics.

An interesting feature of cohorts is that they provide a simple way to study age, cohort, and year effects. This is relevant for our study since informality rates vary dramatically by age and by cohorts. We follow Deaton (1997) and Deaton and Paxson (1993), and consider the case of informality rates. Indexing cohorts with \( c \) (where the index \( c \) is defined as the age of cohort \( c \) in the first year of the data, i.e., 1985), time (years) with \( t \), and age, defined as the age of cohort \( c \) in year \( t \), with \( a \), we can write the informality rate as

\[
i_{ct} = \beta + \alpha_a + \gamma_c + \psi_t + \mu_{ct}.
\]

This provides a natural decomposition of informality rates of cohort \( c \) at period \( t \) into three determinants: an age effect, a cohort effect and a temporal effect. In order to estimate (1), we need to give some structure to each of the effects. If there is no obvious pattern \textit{a priori}, we can use dummy variables, as suggested by Deaton (1997), and let the data reflect any pattern.

\footnote{Static cohort based approaches trace back to the pioneer study by Deaton and Paxson (1992). Dynamic analysis of cohorts is a much recent issue related to recent papers by Collado, McKenzie and Verbeek and Vella. Dynamic cohort methods have been recently used successfully in the analysis of income mobility, like for example Antman and Mackenzie (2005) for the case of Mexico, and Navarro (2006) for Argentina}
Let $A$ be a matrix of age dummies, $C$ a matrix of cohort dummies, and $Y$ a matrix of year dummies. Arranging the data in cohort-year pairs, equation (1) can be rewritten as

$$i = \beta + A\alpha + C\gamma + Y\psi + \mu, \quad (2)$$

where $i$ is the stacked vector of informality rate observations in each cohort-year pair. Even after dropping one column from each of the three matrices, the model is still overparameterized. This is easily seen from the fact that

$$a_{ct} = c + t. \quad (3)$$

Because of this linear dependency between age, cohort, and year, we need one more normalization. First, note that (3) implies

$$As_{na} = Cs_{nc} + Ys_{ny} \quad (4)$$

where $s_n$ is the vector $\{0, 1, 2, 3, \ldots, n\}$. Then, note that (2) can be rewritten as

$$i = \beta + A\bar{\alpha} + C\bar{\gamma} + Y\bar{\psi} + \mu, \quad (5)$$

where, for any non-zero scalar $\kappa$,

$$\bar{\alpha} = \alpha + \kappa s_{na}, \quad \bar{\gamma} = \gamma - \kappa s_{nc}, \quad \bar{\psi} = \psi - \kappa s_{ny}. \quad (6)$$

Therefore, any trend in the data can be arbitrarily reinterpreted as a year trend, or as equal trends in ages and cohorts but of opposite sign (since year equals age minus cohort). Now, one suitable normalization is to attribute the time trend to the age and cohort effects, and let the year effects capture cyclical fluctuations that average zero over the long-run. We can do this by imposing the restriction that the year dummies are orthogonal to a time trend, that is,

$$s'_{ny} \psi = 0, \quad (7)$$

and to add to zero. This is the normalization adopted in the empirical part of the paper.
3 Panel VAR and informality

Not without significant controversy, since the seminal work by Sims (1980) VAR methods have played a leading role in the analysis of dynamic interrelations among macroeconomic variables, specially in cases where the underlying theory is weak or silent about the distinction between endogenous and exogenous variables. In such cases, VAR’s provide a flexible statistical system where dynamic relations can be learned and quantified based on a minimal theoretical structure. This seems to be the case of the relationship between informality rates and other labor market variates, like relative wages, hours worked or unemployment.

The debate on whether informal workers are rationed out or voluntarily accept to enter the sector indicate the coexistence of competing causal theories for these dynamic interrelations. Specifically, the segmentation view, by which informal workers are pushed towards informality, would predict that changes in relative wages follow rather than precede movements in the informal salaried rate, and that the latter comoves with the unemployment rate. Thus, higher unemployment would tend to induce individuals to accept lower pay as informal salaried, and exogenous reductions in informality (say due to stricter enforcement) could increase unemployment. On the contrary, the integrated or voluntary view postulates that workers are drawn into self-employment by attractive movements in the earnings in independent activities relative to formal salaried employment, so that changes in relative earnings precede increases in self-employment. Moreover, the segmentation view predicts that informal employment is countercyclical, consistent with the sector functioning as a safety net during economic deceleration or crisis, while in the integrated view it may behave pro-cyclically if growth is driven by booms in non-tradable sectors which have a higher propensity to operate informally. These distinct predictions are clearly testable with the right data at hand.

Like all time-series based methods, VAR’s adequacy is severely affected by the absence of long series in the appropriate frequency. This is even more relevant when assessing and modeling the non-stationarities induced by unit-root
process since standard tests have serious troubles distinguishing their null and alternative hypotheses when the number of time periods is short. Panel data can alleviate this problem by increasing the number of degrees of freedom through the study of time series of different units of analysis (countries, individuals). Within this literature, the unit-root panel tests of Levin, Lin, Maddala and, or In, Pesaran, provide panel alternatives to standard procedures like the classic Dickey Fuller based tests. As clearly stressed by Maddala, the increased number of degrees of freedom introduced by the cross sectional variability comes at the price of introducing an additional source of heterogeneity. In fact, only under very stringent homogeneity assumptions panel methods can be seen as providing an improvement over standard time series methods.

Heterogeneities, in the form of unobserved random or fixed individual specific terms, have been a major concern in the dynamic panel literature since these render standard panel methods like fixed and random effects estimators inconsistent. This problem has started a large body of theoretical as well as empirical literature starting with the instrumental variable based works of Anderson and Hsiao or Arellano and Bond, to more recent likelihood based approach of Hsiao, to name a few. Arellano (2004) is an authoritative reference on the subject.

All these concerns affect the analysis of PVAR’s since they are in essence a system of dynamic panel equations. The literature on PVAR’s refers back to the paper by Holtz-Eakin who established a framework for identification, estimation and inference. At this point there is not a well established theory where all the previous issues (in particular heterogeneities and non-stationarities) are dealt in a conclusive manner, even though the recent article by Blinder, etc. makes considerable contributions. Thus, the advantages and disadvantages of this method should be assessed on each particular application.

Consider the PVAR(1) specification:

\[ w_{it} = \mu_i + \Phi w_{i,t-1} + \epsilon_{it}, \quad i = 1, \ldots, N, \ t = 1, \ldots, T \]

where \( w_{it} \) is a vector of \( m \) random variables, \( \Phi \) is an \( m \times m \) matrix of coefficients, \( \mu_i \) is a vector of \( m \) individual effects and \( \epsilon_{it} \) is a multivariate white-noise vector.
of $m$ residuals. As with standard VAR models, all variables depend on the past of all variables in the system, the main difference being the presence of the individual specific terms $\mu_i$.

Not being a structural form, VAR’s usefulness comes from implementing dynamic simulations, once the unknown parameters are estimated. For example, impulse-response analysis or variance decompositions are standard exercises. This requires to solve a delicate identification problem, which is usually handled through establishing exogenously a precedence ordering so more exogenous variables impact on the more endogenous ones in a sequential order. This is the standard identification strategy implicit in the Choleski decomposition which induces a recursive orthogonal structure on the structure of the shocks $\epsilon_{it}$. In this paper we make the plausible assumption that informality and unemployment rates are ‘relatively’ more exogenous than relative earnings. However, as discussed below, the particular ordering adopted is irrelevant given the low estimated covariances between the errors across equations.

Regarding estimation and inference, we will use a system-based GMM estimator for each equation, as in Arellano and Bover (1985)\footnote{Estimation is implemented with the PVAR routine by Inessa Love. See Love (2003) for computational details, and Love and Ichino (2006) for a recent empirical application.}

## 4 Data and construction of cohorts

The source of households data is the EPH (Encuesta Permanente de Hogares). This survey is conducted by INDEC (Instituto Nacional de Estadisticas y Censos) since 1972. Until 2003, data was collected twice a year (May and October) with a rolling panel structure. Since 2003, data is collected quarterly.

We use successive years of cross-sectional EPHs to follow cohorts of individuals over time. We classify individuals with ages ranging from 15 to 64 into cohorts according to their year of birth. In order to get the longest continuous time span of surveys, we use data from Great Buenos Aires area only, which allows to have data from 1985 to 2003. We also construct cohorts based on two skill groups: unskilled (incomplete secondary education or less) and skilled
(those with at least a high school diploma). This is important since compositions effects are present, especially in the group of independent workers, which includes both high skilled professionals and low educated individuals with precarious employment. Then, for each cohort-survey year pair we compute three variables: informality rates, unemployment rates, and relative earnings. The educational split is only used for the cohort-year-age decomposition analysis and not in the PVAR given that there is a significant loss of precision as the number of observations in each cell becomes small. Moreover, the results from this analysis strongly suggest very similar trends for both skill groups.

The informality rate is the ratio of informal workers over the total number of employed individuals, and is computed under two alternative definitions. In the first definition, labeled as *informal salaried* we consider as informal all salaried workers without social security contributions in their current job, and to obtain the informality rate this is divided by total employment. The second definition will be referred to as *informal independent*, and treats as ‘informal’ all self-employed (both professional and non-professionals) individuals and the owners of micro firms (those employing less than 6 workers). The unemployment rate is computed as the share of unemployed individuals in the labor force using the official definition. Lastly, relative wage refers to ratio of the mean earnings in informal jobs (for each definition) to the mean earnings in formal jobs.

## 5 Estimation results

We start discussing year, age and cohort effects as described in Section 2. In order to handle the identification problem of cohort effects, we proceed as described before, by estimating (2) subject to (7), for each of our three variables: informality rates (both definitions), relative wages (in both definitions) and unemployment rates. Again, we also split the sample in two groups with high and low education. Results of the decompositions are presented graphically in Figures 1 to 6 attached.

Regarding time effects, we recall that due to the orthogonality conditions
imposed, they are expected to reflect cyclical behavior. For the informal salaried definition, the time effect on informality rates is rather stable. Larger fluctuations are observed for the low skilled workers, with a strong increase after 2000, the period where the most severe part of the crisis started. The case of independent workers markedly different. The temporal effect of informality rates follows a clear counter cyclical pattern and is very similar across individuals with different skills. Starting in 1985 the effect is markedly increasing with a peak in 1991. This is coincident with a period of very high inflation that climaxed in the hyperinflation episode of 1989, and the failed stabilization plan of 1991. In parallel with the adjustment of macroeconomic variables and the implementation of the Convertibility Plan and the sustained growth thereafter, the effect decreases to levels similar (and even lower) to those of 1985. The strong depression that started around 1997 and that eventually led to abandoning the convertibility scheme and a drastic recession was accompanied by an increase in informality according to this definition, though interestingly at a lower pace than the one implied by the hyperinflation crisis in the late eighties.

These results are compatible with the idea that informality may arise as a response to the noise environment posed by high inflation and relative price dispersion and, as suggested by Bour and Susmel (2001) as a more flexible way to negotiate real wages when nominal prices (and maybe wages) move up. This may explain why the increase in informality during the hyperinflation crises was much larger than the one observed during the recession (with stable prices) of the nineties.

Interestingly, the time effect on informal salaried rates does not increase with the hyperinflation episode of the late 80’s, and the cyclical pattern observed for all independent workers appears only when the low-skilled informal salaried are considered: a strong negative effect in the early 90’s, while the country was growing fast, and a sudden increase after the Tequila crisis in 1996. In sum, the counter cyclical nature of these effects goes against the view that informal salaried employment is positively associated to episodes of economic expansion.

Next, consider time effects on relative wages. For the informal salaried
workers the time effect is mildly cyclical, with a peak around 1994, coinciding
with a period of rapid economic growth, whereas for independent workers the
pattern is more stable in its trend and more erratic. It is important to recall
that these effects are fluctuations around the trends captured by cohort and age
effects.

Cohort effects are expected to capture trends. For informal salaried individ-
uals the cohort effect on informality rates is monotonically decreasing while for
the independent workers is strictly increasing, and except for the levels, there
are no differences in the evolutions for different skill groups. Younger cohorts
present a higher incidence of independent work and less informal salaried work-
ers. That is, more recent cohorts are more prone to be employed as informal
salaried and less likely to work independently, which reflect the trends observed
when examining aggregate informality rates in Argentina during this period.

Relative wages present a slight upward trend in both definitions: younger
cohorts faced lower informal to formal relative wages, which again corroborate
the decline in informal to formal earnings that is well established fact in the
literature (Bour and Susmel 2000). Differences across skill groups are small in
the case of informal salaried workers, and very marked in the case of independent
workers. This reflects the fact that the second definition (independent workers)
masks a more heterogeneous composition. Informal salaried workers are, overall,
low salary individuals in more precarious jobs, whereas the group of independent
workers include highly paid professionals as well low skilled individuals.

The cohort effects on unemployment is clearly decreasing. Younger cohorts
were exposed to higher unemployment. Again, this is compatible with well
known upward trend in the unemployment rate experienced by Argentina in
the 1990s.

Finally, age effects reveal interesting patterns, which are somewhat at odds
with cross-section life-cycle employment profiles. The effect for the informal
salaried are decreasing first and markedly increasing after 20 years old. This
is likely the consequence of a composition effect. Those in the labor force that
work at very early ages (15) are mostly disadvantaged individuals or obviously
lack a track record and hence are more likely to accept precarious salaried jobs or hold these as first jobs to accumulate experience. In the extreme, had we considered younger individuals the informality rate would be dramatically high since child labor is illegal. Close to 18 to 19 years there is a sudden increase in the participation rates consistent with the fact that many individuals finish high school and achieve the legal age required by many formal jobs (18 years old). This likely explains the turning point at lower ages. Beyond this point informality rates increase monotonically with age. The case of independent workers is more compatible with recent informality profiles in the sense that informality has a peak around 30-35 years old, with a reversion around 60 years old, which may be related to the proximity of retirement ages. Finally, relative wages decrease with age under both definitions, except for low skilled individuals, for whom it is relatively constant.

Now we turn to the PVAR results shown in Table 1 attached. To preserve degrees of freedom we used a simple PVAR(1) for informality rates, relative wages and unemployment. We performed the analysis for both definitions of informality separately.

Neither for the informal salaried nor the independent workers definition of informality lagged relative wages are significant in the informality rate equation. In the Granger causality framework this means that relative wages do not cause informality rates. Perceived movements in relative wages do not help predict movements in or out of informality, no matter what definition of informality is adopted. This is largely inconsistent with the idea that increases in informality in certain periods is driven by increases in the informality premium. On the other hand, exogenous shocks to informality rate push informal relative salaries down: informality rates Granger-cause relative wages, inducing changes with an opposite sign. Exogenous shocks that push workers into the informal sector are followed by decreases in their relative wage.

A revealing finding pertains to the role of unemployment rates. These Granger-cause informality using the ‘informal salaried’ definition but are uncorrelated with changes in the share of independent work. Exogenous shocks
that increase unemployment induce workers to accept ‘informal salaries’ but do not push people to work independently. Moreover, informality rates and unemployment have a negative impact (in the Granger sense) on relative wages, for the A and the B definition. Exogenous shocks in unemployment or informality rates push informal wages down. Additionally, informality rates in the first definition Granger/cause unemployment whereas the second cases does not. This means that informal salaried employment and unemployment behave as very similar phenomena. This is compatible with the idea that informal salaried employment correspond to inferior jobs and hence its behavior comoves with unemployment. Note that this is consistent with the informal salaried sector being a place holder sector for the pool of potentially unemployed. In contrast, self-employment shows a markedly different, separate dynamic from that of unemployment.

Another important finding is related to the persistence of variables. Informality according to the second definition (independent workers) is much more persistent than in the case of informal salaried employment. This is revealed by the coefficient of lagged informality in the informality equations (Table 1). This is compatible with the idea that being an informal salaried is a relatively more temporary situation (these jobs show high turn over), which workers leave once they find an opportunity to become formal salaried or independent. Meanwhile the status of being independent is more stable since workers tend to enter self-employment at later ages once they have accumulated the required capital and skills. Thus, once they become self-employed, workers tend to remain so, consistent with this being largely a voluntary choice.

Finally, the variance decomposition analysis in Table 1 suggests that exogenous shocks to informality rates (uncorrelated with shocks to unemployment and relative wages) explain almost all its variability. On the other hand, 55% of the variability in relative wages is due to exogenous shocks in informality rates. This again is consistent with the labor market adjusting partially to changes in the relative supply of different types of jobs.

Impulse response analysis provides an alternative representation of all these
effects. Results are shown graphically in Figure 4. An exogenous shock in the informal rate is followed by a negative short run impact on the relative wage premia of the informal sector. It should be noted that variance decompositions and impulse response analysis rely on a particular orthogonalization. The one adopted in the estimation is that informality rates are more exogenous than relative wages. The fact that the correlation across errors of both equations in the PVAR system is 0.0765 suggests that the reduced form errors used in the PVAR specification are uncorrelated hence the particular ordering in the Choleski decomposition should be irrelevant to the results. However, Granger causation is essentially temporal precedence and hence care must be taken to make true causal inferences.

6 Concluding remarks

The interrelation between informality and the wage premia for being in the informal sector takes place in a highly interactive context where empirical evidence and existing theories are not conclusive about mutual interrelations. In particular, informality may arise as a response to bad economic conditions with a consequent impact on relative wages, or as a way to take advantages of more flexible and maybe higher salaries in the informal sector. In order to study these interrelations, this paper implements a panel VAR approach to study the dynamics of informality and key labor market variables like relative wages and unemployment rates. The very short time span of panels based on household level data is overcome by constructing cohorts of individuals that can be followed over long period. Besides, the aggregation process implicit in the construction of cohort level variables helps mitigate measurement errors present at the individual level. We know of no other study that uses this approach for the analysis of informality.

The paper yields novel and interesting evidence on the two main hypotheses of informality. Overall, the results are consistent with the idea of the exclusion view of informal salaried employment as lower quality jobs, while independent
employment is largely driven by voluntary choice of workers (that is persistent) although with a minor subgroup of workers that may be excluded from more desirable formal salaried jobs (that moves with the economy cycle).

This is supported by the following findings:

i) Both informal salaried and independent employment are countercyclical. Moreover, independent employment tends to show larger swings under periods of high price instability. This is compatible with the idea that during economic expansions the opportunity cost of informal is higher, and that during noisy price environment independent employment offers more flexibility to accommodate real earnings movements. ii) Pvar results indicate that informal salaried and unemployment mutually cause each other (positively) in a Granger sense, but that this is not true for unemployment and independent employment. This means that exogenous shocks that induce unemployment (informal salaried employment) are followed by increases in informal salaried employment (unemployment). In contrast, exogenous shocks to self-employment do not help predict unemployment. iii) Pvar results show that both informal salaried work and independent employment Granger-cause relative formal-informal earnings, negatively, that is shifts in the relative shares of either kind of informal employment are followed by declines in their relative earnings (and the converse). Quite strikingly, exogenous shocks to relative formal-informal earnings do not help predict movements in and out informality. This is consistent with the push hypothesis of what drives informality, rather than the pull view which sees workers being drawn into informal employment by better earnings prospects. iv) However, in the case of independent work, the findings of countercyclical behavior and lack of (Granger) causality from relative earnings to self-employment are consistent with: a) the involuntary component of this sector expanding or shrinking with the business cycle while, as theory predicts, the voluntary component can not unambiguously be signed in the cycle (since it depends on the type of boom on tradable led expansions the opportunity cost of being self-employed tends to go up while on non-tradable booms it goes down); b) individuals with entrepreneur qualities may be drawn into self-employment attracted by good business oppor-
tunities but in equilibrium the average relative earnings in the sector need to go down as more individuals enter the sector.

v) Pvar results show that independent work is significantly more persistent than informal salaried employment, that is, the former seems a relatively more permanent state (consistent with idiosyncratic motives) than the latter (consistent with lower tenures).

Thus, in the case of informal salaried workers the results are consistent with the exclusion hypothesis by which these workers are pushed towards informality since they lack better employment options. Higher unemployment would tend to induce individuals to accept lower pay as informal salaried, and exogenous reductions in informality (say due to stricter enforcement) could increase unemployment. These results are consistent with the idea that informal salaried employment is mainly a result of firm decisions and not a workers choice. This may reflect labor rigidities that segment the labor market or implicit collusion to evade income, payroll and other taxes.

In the case of independent workers, the evidence suggests a two tier sector. A majority component seems to enter voluntarily and stay there longer, although there is no evidence that workers are drawn into self-employment by attractive movements in the earnings in independent activities relative to formal salaried employment. However, a second component behaves countercyclically, consistent with the sector also serving as a safety net during economic deceleration or crisis.

In light of the scarcity of time-series based studies of informality at the micro level, the results of this paper offer new light to the debate of the nature of informal employment in developing countries like Argentina. Future studies for other countries could corroborate the findings and subject to more detailed robustness analysis. In particular, a larger VAR system with richer dynamics is a natural extension, as well as exploring whether panel regression results differ when the sample is split into possibly heterogeneous groups like individuals with high and low education.
7 References


Binder, M., Cheng Hsiao and M. Hashem Pesaran, 2000, Estimation and Inference In Short Panel Vector Autoregressions with Unit Roots And Cointegration.


Fiess, N. M. Fugazza and W. Maloney, 2002, Exchange Rate Appreciations,


### TABLE 1. PVAR(1) estimates: Informality rates and relative wages

#### A Definition

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>se</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Informality Rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lir</td>
<td>0.65194</td>
<td>0.18888</td>
<td>3.45166</td>
</tr>
<tr>
<td>Lrw</td>
<td>0.03423</td>
<td>0.03962</td>
<td>0.86400</td>
</tr>
<tr>
<td>Lur</td>
<td>0.58203</td>
<td>0.08827</td>
<td>6.59363</td>
</tr>
<tr>
<td><strong>Relative Wage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lir</td>
<td>-3.27878</td>
<td>1.72183</td>
<td>-1.90424</td>
</tr>
<tr>
<td>Lrw</td>
<td>-0.67007</td>
<td>0.42447</td>
<td>-1.57859</td>
</tr>
<tr>
<td>Lur</td>
<td>-2.04342</td>
<td>0.92709</td>
<td>-2.20413</td>
</tr>
<tr>
<td><strong>Unemployment Rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lir</td>
<td>0.57802</td>
<td>0.25112</td>
<td>2.30178</td>
</tr>
<tr>
<td>Lrw</td>
<td>0.08543</td>
<td>0.06040</td>
<td>1.41446</td>
</tr>
<tr>
<td>Lur</td>
<td>0.78814</td>
<td>0.12167</td>
<td>6.47778</td>
</tr>
</tbody>
</table>

#### Variance decompositions

<table>
<thead>
<tr>
<th></th>
<th>s</th>
<th>inwb</th>
<th>nwrb</th>
<th>urwb</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Informality Rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.5791</td>
<td>0.0148</td>
<td>0.4061</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.5564</td>
<td>0.0155</td>
<td>0.4281</td>
<td></td>
</tr>
<tr>
<td><strong>Relative Wage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.3749</td>
<td>0.3513</td>
<td>0.2738</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.5302</td>
<td>0.0604</td>
<td>0.4095</td>
<td></td>
</tr>
<tr>
<td><strong>Unemployment Rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.5202</td>
<td>0.0400</td>
<td>0.4399</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.5508</td>
<td>0.0179</td>
<td>0.4313</td>
<td></td>
</tr>
</tbody>
</table>

Sample: individuals between 15 and 64 years old
Informality Definition: (INFORMAL SALARIED RATE) salaried workers lacking contributions to social security in their current job

#### B Definition

PVAR(1) estimates: Informality rates and relative wages

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>se</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Informality Rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lir</td>
<td>0.92875</td>
<td>0.34878</td>
<td>2.66286</td>
</tr>
<tr>
<td>Lrw</td>
<td>0.05882</td>
<td>0.04327</td>
<td>1.35946</td>
</tr>
<tr>
<td>Lur</td>
<td>0.24641</td>
<td>0.26388</td>
<td>0.93382</td>
</tr>
<tr>
<td><strong>Relative Wage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lir</td>
<td>-5.62432</td>
<td>2.72342</td>
<td>-2.06517</td>
</tr>
<tr>
<td>Lrw</td>
<td>-0.63687</td>
<td>0.33390</td>
<td>-1.90740</td>
</tr>
<tr>
<td>Lur</td>
<td>-5.17871</td>
<td>2.08998</td>
<td>-2.47788</td>
</tr>
<tr>
<td><strong>Unemployment Rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lir</td>
<td>0.95003</td>
<td>0.51012</td>
<td>1.86237</td>
</tr>
<tr>
<td>Lrw</td>
<td>0.10538</td>
<td>0.06288</td>
<td>1.67584</td>
</tr>
<tr>
<td>Lur</td>
<td>1.47324</td>
<td>0.38407</td>
<td>3.83585</td>
</tr>
</tbody>
</table>

#### Variance decompositions

<table>
<thead>
<tr>
<th></th>
<th>s</th>
<th>inwb</th>
<th>nwrb</th>
<th>urwb</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Informality Rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.73082</td>
<td>0.00372</td>
<td>0.26546</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.70042</td>
<td>0.00050</td>
<td>0.29908</td>
<td></td>
</tr>
<tr>
<td><strong>Relative Wage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.69493</td>
<td>0.00714</td>
<td>0.29793</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.70017</td>
<td>0.00052</td>
<td>0.29931</td>
<td></td>
</tr>
<tr>
<td><strong>Unemployment Rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.69478</td>
<td>0.00137</td>
<td>0.30385</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.70017</td>
<td>0.00049</td>
<td>0.29934</td>
<td></td>
</tr>
</tbody>
</table>

Sample: individuals between 15 and 64 years old
Informality Definition: (INDEPENDENT WORKERS RATE) self-employed (professionals and non-professionals) and owners of micro firms (<6 employees)
FIGURE 1. Age Effects. A Definition

Informality Rate

Relative Wage

Unemployment
FIGURE 2. Cohort Effects. A Definition

Informality Rate

Relative Wage

Unemployment
FIGURE 3. Year Effects. A Definition

Informality Rate

Relative Wage

Unemployment
FIGURE 4. Age Effects. B Definition

Informality Rate

Relative Wage

Unemployment
FIGURE 5. Cohort Effects. B Definition

Informality Rate

Relative Wages

Unemployment
FIGURE 6. Year Effects. B Definition

Informality Rate

Relative Wages

Unemployment
FIGURE 7. Impulse-Response Functions

A Definition

Impulse-responses for 1 lag VAR of ira rw urb

Errors are 5% on each side generated by Monte-Carlo with 500 reps

B Definition

Impulse-responses for 1 lag VAR of irb rwb urb