Unpacking the Persistence of Informality*

Benjamín Villena-Roldán

Universidad Andres Bello, Institute for Economic Policies;

LM^2C^2; and MIPP

April 17, 2023

Abstract

In emerging economies, policymakers should care not only about the informality level but also about its persistence, which also has key welfare implications. Considering worker flows that drive people into informality and prevent them from leaving is important for developing effective public policies. Using a first-order Markov approximation, I find relatively low persistence of informality in Chile, implying a duration of nearly two quarters, and estimate the contributions of labor market transitions among formality, informality, unemployment, and inactivity. The role of flows into informality from formal employment and inactivity and an indirect outflow from formality to inactivity are the main forces in shaping the persistence of informality. I also tackle an additional challenge of appropriately measuring informality dynamics since worker transitions, particularly informal ones, became much harder to track during the COVID-19 pandemic. Finally, I find that if the composition of the Chilean population was that of 2017, informality persistence would be 26% lower, a change driven to a large extent by an increase in groups with high informality attachment.

Keywords: Informality, transition rates, non-response bias, composition bias.

*Email: benjamin@benjaminvillena.com. The author thanks and acknowledges the financial support from FONDECYT grant project 1191888; ANID Milenio Labor Market Mismatch - Causes and Consequences, LM^2C^2 (NCS2022_045); and the Institute for Research in Market Imperfections and Public Policy, MIPP (ICS13_002 ANID). All potential errors are mine.

JEL Codes: J46, J64, E24

1 Introduction

In emerging economies, a large proportion of the workforce is informal. On top of traditional concerns such as poor working conditions and low productivity, the COVID-19 pandemic showed that governments had a hard time reaching informal workers with subsidies or in-kind transfers to reduce their mobility and assuage the disease's spread.

While the level of informal employment is important, a key understudied topic is whether informality is permanent or transitory. Starting with Maloney (1999) few papers have addressed this issue: Gong et al. (2004), Bosch and Maloney (2010) and Escobedo and Moreno (2020) provide evidence from Latin American countries. Akay and Khamis (2012) also find substantial informality persistence in Ukraine. Kucera and Xenogiani (2009) explore the role of institutions in shaping informality across underdeveloped economies. Günther and Launov (2012) highlight the mobility between formal and informal market segments. If informality is a way to avoid unemployment but generates low living standards, its high persistence becomes a trap for workers that appropriate policies could mitigate. Moreover, as shown by Busso et al. (2021), the COVID-19 lockdowns make it difficult for informal workers to get government transfers, so their vulnerability may worsen as they remain in the informal sector.

The measurement of informality persistence depends on the level of aggregation we use to analyze the data. Here I focus on informality and its related transitions for average workers. Admittedly, average behavior may hide relevant micro-heterogeneity which I leave for future research. To assess the importance of compositional changes on the average informality, I use the weighting technique of DiNardo et al. (1996) (DFL henceforth) to construct counterfactual transitions as if the composition were that of 2017, the year of the unemployment rate closest to the pre-pandemic average. Changes in workforce composition explain trends in informality, as Haanwinckel and Soares (2021) find in Brazilian data, using a structural model approach.

Another relevant empirical challenge is the non-response rate in a rotating panel of employment surveys, which may affect the level and dynamics of variables, as found by Neumark and Kawaguchi (2004). In CPS data (Bernhardt et al., 2021) and comparable Chilean data, this issue had increased even before COVID-19. I use the inverse probability weighting (IPW) technique of Wooldridge (2007) to deal with this problem, which essentially increases the importance of individuals with a low likelihood of being re-interviewed.

Using IPW and IPW-DFL weights, I report labor market stocks and the quarterly transition series for 2011-2021. The evidence shows that informality smoothly declined over this period. This trend was interrupted by a sharp reduction right after the beginning of the COVID-19 pandemic and a subsequent recovery close to previous levels. Worker transition rates also show a mildly decreasing pattern from formal to informal jobs over this period and a smooth increasing rate of formal stayers. The onset of the pandemic generated a huge amount of turmoil with abrupt transitions for all labor market states.

Then, I propose a first-order Markov approximation to decompose informality persistence into contributions of transition rates between formal employment (f), informal employment (i), unemployment (u), and inactivity (o). Using this decomposition, whose components I show I can obtain from a series of OLS regressions, I conclude that the estimated persistence in Chile implies an average informality duration of nearly one semester, substantially lower than estimates available from Argentina, Mexico, and Brazil of a year, as shown by Bosch and Maloney (2010). Even though non-response in the employment survey does not change the level of stocks or transition variables by much, it does lead to an overestimation of informality persistence.

Moreover, I show that an indirect informality outflow, the inactivity-to-formality ($o2f$) quarterly transition, has a large negative association with the persistence of the informality-to-population ratio, suggesting a likely time-aggregation measurement issue with substantial intra-quarter transitions. Informality inflows from formal workers ($f2i$) and from inactive individuals ($o2i$) are also important for understanding the roots of informality persistence. This evidence suggests that policymakers trying to achieve lower informality should aim at preventing workers from entering this state rather than pulling them out of it.

I also use DFL-weighted estimates to assess the role of the compositional change of the Chilean population in shaping informality. If the composition remained as in 2017, the informality persistence would be 26% lower as the shares of immigrants, females, and other groups with a higher attachment to informality increased. However, with fixed composition, the protagonist flows under the case of varying composition remain the same, but other flows such as formality-to-inactivity ($f2o$) and inactivity-to-informality ($o2i$) become more relevant.

The rest of the paper is as follows: in Section 2 I present the data and explain both non-response and composition issues. Then, I show how to construct series while taking them into account. In Section 3 I explain the main setup to decompose stock persistence into worker flow contributions and show how this can be empirically implemented. Section 4 shows the main results. First, I plot and explain stock and flow variables using IPW and IPW-DFL weights. Second, I show the persistence decomposition of informality and discuss its interpretation. Third, I show how different assumptions may affect the decomposition results obtained. Finally, Section 5 presents the conclusions.

2 Data

I study labor transitions using the National Employment Survey (ENE in Spanish) of the Chilean www.ine.cl. Informal workers are those who have a labor contract and contribute to the pension and health systems, regardless of the funding source or whether it is public or private.

The ENE survey has a quarterly rotating panel structure and stratified sampling. Individuals in urban households are interviewed for up to six consecutive quarters before being replaced in the sample. Using these data, I compute the transition probabilities from and to formal employment (f), informal employment (i), unemployment (u), and out-of-the-labor-force (o). Data construction is described in appendix A.1.

Formally, the ENE survey has probability weights from the stratified sampling based on the 2017 National Census, denoted as $\omega_{n,t}$ for an individual n at time t. A stock of individuals in state $x = \{f, i, u, o\}$ at time t is denoted as x_t. Without taking into account the non-response of respondents, it is computed as

$$x_t = \sum_{n=1}^{N_t} \omega_{n,t} x_{n,t}$$

where $x_{n,t}$ is an indicator that individual n is in state x at time t. N_t stands for the number of surveys at time t. These are labeled in the main paper *survey* estimates. As an illustration, the traditional definition of the unemployment rate is computed as $\frac{u_t}{u_t+e_t}$, with $e_t \equiv f_t + i_t$.

Using survey weights, a transition rate from state x to state z completed at time t is computed as

$$x2z_t = \frac{\sum_{n=1}^{N_t} \omega_{n,t} s_{n,t} x_{n,t-1} z_{n,t}}{\sum_{n=1}^{N_t} \omega_{n,t} s_{n,t} x_{n,t-1}}$$

where $s_{n,t}$ is a dummy variable that takes the value 1 if the individual n observed at time $t-1$ remains in the sample at time t, and 0 otherwise. This formulation highlights that transition rates are computed for individuals transiting from state x to z and remaining in the sample for consecutive periods. Weights $\omega_{n,t}$ must be the same to ensure that the numerator and denominator capture the same number of individuals who remain in sample over two periods.

There are two concerns with an unbiased estimation of persistence: changes in the composition of the labor force and the non-response sample selection.

2.1 Composition issues

As individuals differ in their attachment to informality, the composition of the population explains at least part of the persistence. Immigrants, females, college-educated workers, and older workers have largely increased their share of the Chilean labor force. Hence, I use the weighing technique of DiNardo et al. (1996) (DFL henceforth) to keep fixed the composition of 2017, when the unemployment rate (6.9%) is the closest to the pre-COVID-19 average. I estimate the probability of an observation occurring in 2017 in a logit model as a function of categorical variables: sex, immigrant status, head-of-household, marital status, age group, educational group, cohort, and region. In Table A1 in the appendix, I show that while females and heads of households are more frequent in 2017 compared to other years, female heads, singles, and divorcees are less likely than their male counterparts. Chilean females are less represented than their immigrant counterparts, too.

Then, the DFL weight is the predicted probability odds of being in the 2017 sample, i.e.

$$\phi_{n,t} = \frac{q_{n,t}^{2017}}{1 - q_{n,t}^{2017}},$$

where $q_{n,t}^{2017}$ is the corresponding logit-model predicted probability for an individual to be observed in 2017.

2.2 Non-response bias correction

To deal with non-response, I use the IPW method, assuming *unconfoundedness*, as in Wooldridge (2007). In the current context, this means that selection occurs based only on the same observable features of workers as in the previous subsection. Among demographic factors, gender matters since females are more likely to respond to surveys (Groves, 2006). Married and head-of-household individuals respond more in follow-up interviews as they have more stable living arrangements, as do older and more educated workers. Finally, immigrant workers are harder to track down as they move more often.

Considering the effect of these demographics on non-response is important since female labor participation steadily rose over 2010–19 and collapsed as COVID–19 hit. Furthermore, the proportion of the population that was married decreased, and an unprecedented wave of immigration occurred during this time period. IPW weights increase the representation of individuals with a low likelihood of being re-interviewed. In Table A1 in the appendix, I show results along these lines: females, Chileans, household heads, married, and single individuals are more likely to be re-interviewed in the next quarter.

Formally, to deal with the non-response bias, we run logit models using covariates in the ENE survey. Once we have estimated probabilities of response $p_{n,t}$, I compute stocks as

$$x_t^* = \sum_{n=1}^{N_t} \frac{\omega_{n,t}\overline{p}_t}{p_{n,t}} x_{n,t}$$

where $x_{n,t}$ is an indicator that individual n is in state x at time t. N_t stands for the number of surveys at time t. In addition, $\overline{p}_t$ is defined so that the overall sum of weights remains the same as the estimated level of individuals determined by the original sampling design. Therefore, the value $\overline{p}_t$ satisfies

$$\sum_{n=1}^{N_t} \frac{\omega_{n,t}}{p_{n,t}}\overline{p}_t = \sum_{n=1}^{N_t} \omega_{n,t}$$

Hence, $\overline{p}_t$ is an harmonic weighted average of estimated probabilities of response:

$$\overline{p}_t = \frac{\sum_{n=1}^{N_t} \omega_{n,t}}{\sum_{n=1}^{N_t} \frac{\omega_{n,t}}{p_{n,t}}}$$

As a consequence, the non-response corrected probability weights are

$$\omega_{n,t}^{*} = \frac{\omega_{n,t}}{p_{n,t}} \overline{p}_t$$

Then, a generically corrected transition rate between states x and z is

$$x2z_t^{*} = \frac{\sum_{n=1}^{N_t} \omega_{n,t}^{*} s_{n,t} x_{n,t-1} z_{n,t}}{\sum_{n=1}^{N_t} \omega_{n,t}^{*} s_{n,t} x_{n,t-1}}$$

To implement the DFL correction on top of the non-respond adjustment, I just use the weight

$$\omega_{n,t}^{**} = \frac{\omega_{n,t}}{p_{n,t}} \overline{p}_t \frac{\phi_{n,t}}{\overline{\phi}_t}$$

where the last division by the average of weights ensures that the total sum of individuals at time t, defined by the original sample design of the survey, holds.

3 Assessing informality persistence

To assess the sources of informality persistence, as Shimer (2012) and Elsby et al. (2015), I rely on a first-order Markov chain approximation in which stocks at time t, $x_t = \{f_t, i_t, u_t, o_t\}$ stochastically evolve from the previous period with probabilities of transition from state x to state z in the next period, denoted $x2z$. Equation (1) describes the evolution of stocks

$$\begin{pmatrix} f_t \\ i_t \\ u_t \\ o_t \end{pmatrix} = \begin{pmatrix} 1 - i2f - u2f - o2f & i2f & u2f & o2f \\ f2i & 1 - f2i - u2i - o2i & u2i & o2i \\ f2u & i2u & 1 - f2u - i2u - o2u & o2u \\ f2o & i2o & u2o & 1 - f2o - i2o - u2o \end{pmatrix} \begin{pmatrix} f_{t-1} \\ i_{t-1} \\ u_{t-1} \\ o_{t-1} \end{pmatrix} \tag{1}$$

In a steady state, a stock x equals a function of probabilities $x2z$ for $x \in \{f, i, u, o\}$ and

$z \in \{f, i, u, o\} - \{x\}$. Thus, the first-order approximation for stock y at time t is

$$y_t \approx y(\bar{p}) + \sum_x \sum_{z \neq x} \frac{\partial y}{\partial x2z}(\overrightarrow{x2z}) \left(x2z_t - \overline{x2z}\right)$$

where $\overline{x2z}$ is the average transition probability from state x to z, and $\overrightarrow{x2z}$ is a vector containing all average transition probabilities.

To assess the average persistence of a stock, I take the derivative with respect to the lag value of y and average over all periods $t = 1, 2, ..., T$ to obtain

$$\frac{1}{T} \sum_t \frac{\partial y_t}{\partial y_{t-1}} \approx \frac{1}{T} \sum_t \sum_x \sum_{z \neq x} \frac{\partial y_t}{\partial (x2z)_t} \frac{\partial (x2z)_t}{\partial y_{t-1}} \equiv \delta_y \tag{2}$$

The average persistence of a stock y depends on (i) the effects of transition probabilities on the stock (a contemporaneous effect) and (ii) the effects of the previous state y on transition probabilities (a dynamic effect). Equation (2) decomposes persistence into the contributions of each transition.

The first component $\alpha_y^{xz} = \frac{\partial y}{\partial (x2z)}$ is estimated by the regression (3), in line with a standard interpretation of OLS coefficients as average partial derivatives. The stock y is the dependent variable, and the transition rates $x2z$ are the regressors of interest. I also include other controls: quarterly seasonal dummies, a linear trend, a post-2019 dummy, and its interaction with a linear trend. The latter variables intend to capture a likely structural break in the labor market due to fear of contagion and lockdowns generated by the COVID-19 pandemic. The term ϵ_t represents a random error component.

$$y_t = \alpha_0 + \sum_x \sum_{z \neq x} \alpha_y^{xz} (x2z)_t + \text{ other controls } + \epsilon_t \tag{3}$$

The second component $\beta_y^{xz} = \frac{\partial x2z_t}{\partial y_{t-1}}$ is estimated via OLS in (4). The rationale guiding the linear regression in this equation is the one justifying equation (3). Since we assume the data is generated by a first-order Markov chain, controlling for lagged transition rates is needed for the estimated coefficient to be an average partial derivative, i.e.

$$x2z_t = \beta_0 + \beta_y^{xz} y_{t-1} + \sum_r \sum_{q \neq r} \gamma_y^{rq} (r2q)_{t-1} + \text{ other controls } + \upsilon_t \tag{4}$$

where υ_t represents a random error component and $r \in \{f, i, u, o\}$ and $q \in \{f, i, u, o\} - \{r\}$. The total contribution of the flow $x2z$ is, according to equation (2), the multiplication $\alpha^{xz}\beta^{xz}$. A positive α^{xz} indicates that the flow $x2z$ contemporaneously increases the stock y, which may occur even if $x2z$ is not a direct flow in or out of y. As the ENE survey allows us to measure quarterly transitions, an effect from an indirect transition likely reflects that real flows are taking place at a frequency higher than a quarter, i.e., there is some time aggregation bias that could be corrected either in continuous time Shimer (2012, 2013) or in discrete time Elsby et al. (2009); Choi et al. (2015) under the assumption that transition rates remain constant within the quarter. Unfortunately, this low frequency of observation is part of the sampling design of the survey and cannot be changed, making the mentioned assumption unlikely since search effort, incentives, and seasonal factors may vary within a sufficiently long period, such as a quarter. Thus, I rely on feasible measures without making far-fetched assumptions.

The percentage contribution of each transition to the persistence level is computed as

$$S_y^{xz} = \frac{\alpha_y^{xz}\beta_y^{xz}}{\sum_q \sum_{r \neq q} \alpha_y^{qr} \beta_y^{qr}} \tag{5}$$

4 Results

4.1 Evolution of stocks and flows

Figure 1 depicts series for the adult population using survey weights, IPW, and joint IPW-DFL weights. Survey and IPW series look similar. The informal-to-employed ratio (i/e) declines until 2014, hovers around 38% until the sharp drop of the COVID-19 pandemic, and then increases, reaching a 35% in 2021Q4. The informal-to-population ratio (i/a) exhibits a similar pattern. As in many countries, mobility dramatically went down soon after the first official case of COVID-19 was detected on March 3, 2020 (Gozzi et al., 2021). The Chilean health authorities started municipality-level lockdowns very soon based on the number of contagions reported (Bennett, 2021). These events clearly disrupted the labor market's functioning in 2020Q1, generating sudden changes in stocks and flows in that quarter.

The DFL line shows that if we kept the composition of 2017 fixed, the i/e ratio would have been larger in 2010–14 and the i/a ratio would have been smaller. Thus, individuals entering the population from 2017 on are more likely to work and to work informally than incumbents.

The unemployment rate in the lower-left panel falls until 2014, then steadily climbs, sharply increases when COVID-19 hits, and rapidly declines to reach near 8%. DFL weights do not matter in this case. The lower-right panel shows a steadily declining inactivity-to-population ratio (o/a), until a large increase in 2020Q1, which partially reverts to the pre-pandemic level. The DFL line reveals that individuals entering the population after 2017 have a larger attachment to the labor force.

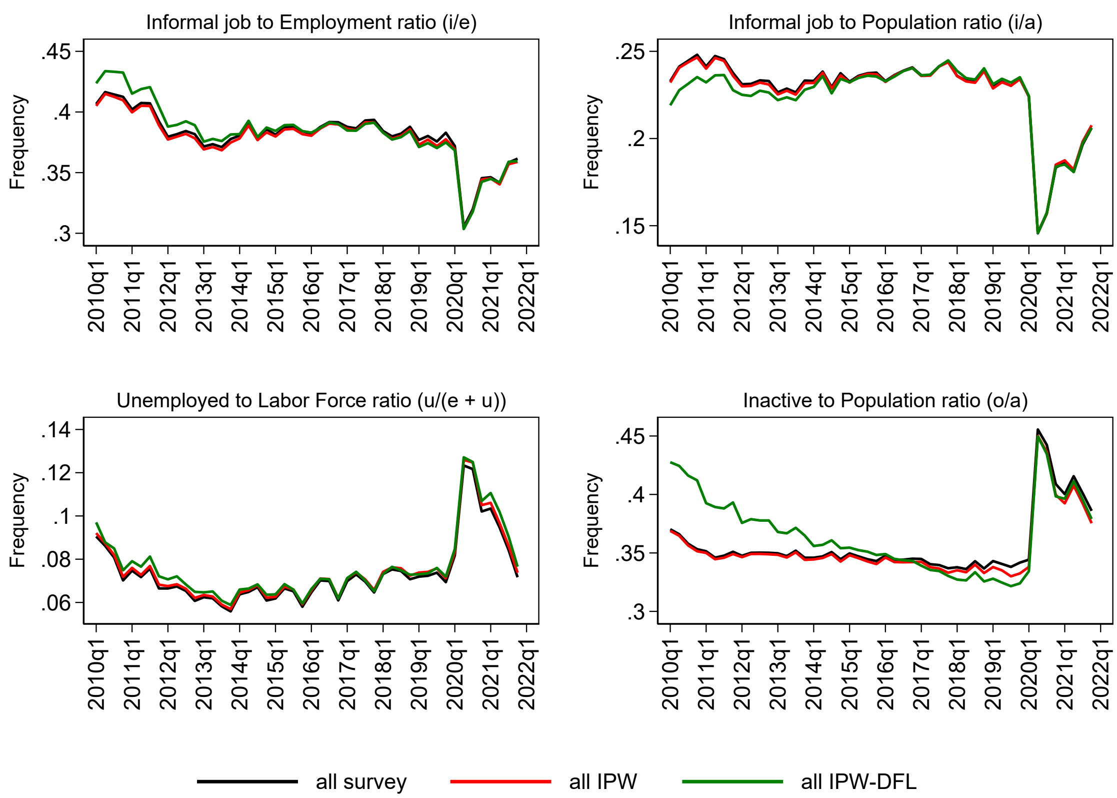

Figure 1: Evolution of informality, unemployment and inactivity ratios

Figure 2 shows informality outflows. At a first glance, neither IPW nor DFL adjustments appear to affect the big picture: after a decade of a mild growing perpetuation of informality, i.e., a soft $i2i$ rise, at the expense of a mild declining "formalization" ($i2f$) and "inactivation" ($i2o$) of informality, the pandemic generates a great escape from informality. The sharp drop of informality-stayers ($i2i$) mirrors sudden spikes in transitions from informality to formality (nearly a threefold increase), unemployment (twofold), and inactivity (in between the last two). After 2020Q1, there is a quick reversal for the most part, but the $i2f$ flow sank below the pre-pandemic level and started rising again until the end of the sample in 2021Q4. Moreover, the informality-stayer flow reached levels somewhat below the pre-pandemic average. Informality flows to both unemployment and inactivity, after the big reversion, stayed at rates above their pre-pandemic average.

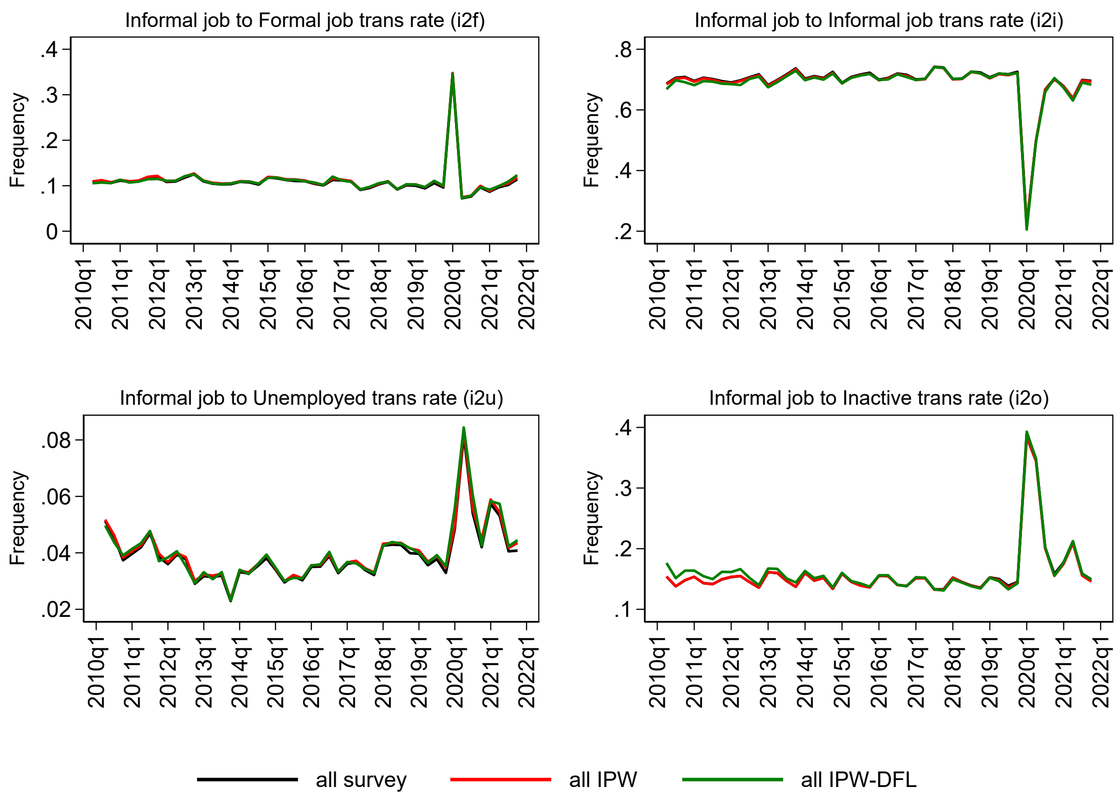

Figure 2: Evolution of informality outflows transition rates

A mildly increasing rate of workers staying at formal jobs ($f2f$) more than halved in the onset of the pandemic, reaching a trough below 30%, according to Figure 3. In contrast, the "informalization" flow ($f2i$) shows a softly declining pattern, until the sudden spike in

2020Q1, as do both transitions from formality to unemployment ($f2u$) and inactivity ($f2o$). The reversion after the pandemic led $f2i$ and $f2u$ below pre-pandemic rates, mirroring the trajectory of $f2o$.

Figure 3: Evolution of formality outflows transition rates

Figure 4 depicts relatively stable trajectories of unemployment outflows to both formal ($u2f$) and informal ($u2i$) employment during 2010–19. Both upper panels suggest that 20-25% of unemployed workers ended up after a quarter with either a formal or an informal job. There is also an upward trend of unemployment stayers $u2u$ and a declining transition to inactivity $u2o$. In the pandemic's onset, unemployed workers suddenly escaped to formality and inactivity but shied away from informality. The aftermath of the pandemic shows a notably lower transition from unemployment to formal employment, which contrasts with somewhat higher flows to either staying in unemployment or becoming inactive.

Quite stable inactivity outflows are shown in Figure 5 until 2020Q1, with just a slight upward trend in transitions to unemployment since 2013. When COVID-19 hit first, there

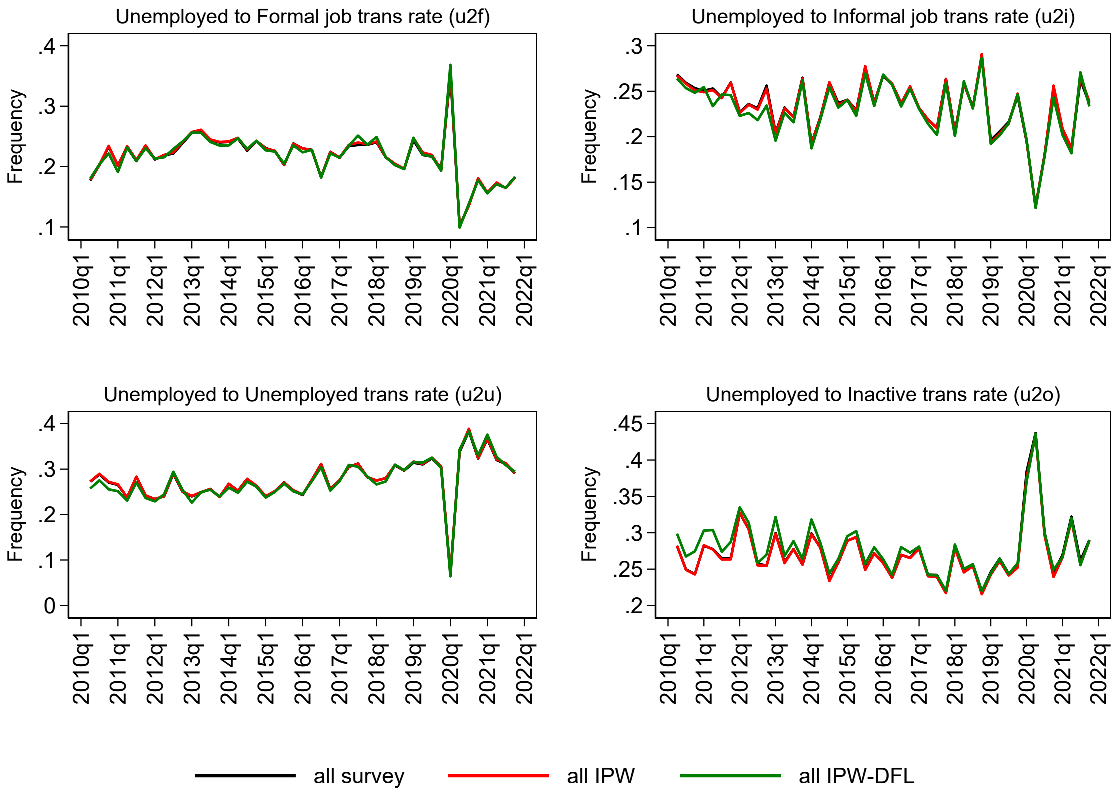

Figure 4: Evolution of unemployment outflows transition rates

was a sudden and sharp decline of inactivity stayers (from 85% to 40%), which mostly mirrored large spikes in flows into both formal ($o2f$) and informal ($o2i$) employment. After 2020Q1, flows quickly bounced back to numbers close to their pre-pandemic levels, but inactivity-stayers remained higher and outflows to informality were lower.

For the sake of completeness, Table A2 in the appendix contain descriptive statistics of all stocks and flows under different weighting schemes.

4.2 Informality persistence decomposition

Table 1 shows the main results of the decomposition for the i/a ratio, controlling for COVID-19. The quarterly persistence is $\delta = 0.515$, implying $1/(1 - 0.515) \approx 2.1$ quarters of average informality duration. If the 2017 composition remained fixed, the figure would be 26% lower, with $\delta = 0.412$, implying a mean duration of 1.7 quarters, a result explained by a larger share of more likely informal individuals, such as females and immigrants. The columns labeled α

Figure 5: Evolution of inactivity outflows transition rates

and β contain the parameters in equations (3) and (4), respectively, and the column labeled $\alpha\beta$ is just the multiplication. Column S shows the share of persistence attributable to each flow, as stated in (5). The $o2f$ flow – inactive people becoming formal workers within a quarter – has the largest absolute contribution, accounting for $S = -110\%$ of all informality persistence. In this case, $\alpha = -0.733$, showing a negative contemporaneous association between this flow and the informality-to-population rate. For $o2f$, $\beta = 0.776$ implies that previous informality is associated with a larger transition from inactivity to formality. The importance of this flow, whose relevance is unexpected at least at higher frequencies of data, suggests that there are substantial intra-quarter short-term flows that make this channel relevant at the quarterly frequency and/or misclassification errors.

The two most important positive contributions to informality persistence are $f2i$ and $o2i$ transitions. Both flows into formality can account for a $S = 93\%$ and $S = 75\%$, respectively. Both inflows show a large positive contemporaneous association to the informality-to-population ratio and a somewhat weaker link to the lagged i/a ratio.

Hence, informality's perpetuation is, to a great extent, a matter of attraction to this state linked to larger inflows from formality and inactivity. Perhaps surprisingly, direct outflows from informality to other states do not play a big role. Instead, an indirect route from inactivity to formality is the dominant driver.

Under compositional adjustments, i.e. keeping the composition as it was in 2017 via DFL adjustment, the same big forces seem to account for informality persistence, especially the $o2f$ flow, but their relevance is even stronger. The importance of direct inflows $f2i$ and $o2i$ also increases their importance as main drivers. A constant composition of the population also gives higher protagonism to the $f2o$ flow that positively affects informality persistence, showing that inactivity is a state in which individuals become informal more easily than do formal workers. Also, the flow from informality to inactivity, $i2o$ becomes an important factor in decreasing informality persistence.

Considering both decompositions in Table 1, I conclude that groups less represented in 2017 compared to other periods can trace their persistence mostly to flows into informality. If policymakers want to reduce the informality rate, more effort should be put into preventing individuals from landing in this state rather than pulling them out of it. In all cases, to understand the transitional dynamics that lead to longer informality spells (higher persistence), it is very important to take into account transitions from and to inactivity, an issue likely belittled in policy discussions.

4.3 Further analysis

Table 2 computes the same decomposition of inactivity persistence using just survey data, i.e., without non-response correction via IPW. Failing to consider this adjustment slightly increases the estimation of overall informality's persistence to $\delta = 0.535$. The main flows accounting for the informality's persistence are the same, even though their relative importance seems greater in this case. In the right-side of the Table, I also report the DFL-adjusted decomposition, showing larger impacts overall for the flows that are relevant for the case in which composition actually varies. Hence, although the broad picture remains qualitative, when non-response is neglected in the analysis, informality seems more reactive to various

Table 1: Decomposition of informality-population ratio persistence, IPW

	no DFL adjustment				DFL adjustment			
transition	α	β	$\alpha\beta$	S	α	β	$\alpha\beta$	S
f2i	0.875	0.548	0.480	0.932	0.783	0.934	0.731	1.773
f2u	0.438	0.211	0.093	0.180	0.294	0.352	0.104	0.251
f2o	0.267	0.916	0.244	0.475	0.332	1.842	0.611	1.481
i2f	0.153	0.369	0.056	0.109	0.083	1.106	0.092	0.223
i2u	0.162	0.180	0.029	0.057	-0.012	0.264	-0.003	-0.008
i2o	-0.265	0.329	-0.087	-0.169	-0.304	1.169	-0.355	-0.861
u2f	0.009	-0.085	-0.001	-0.002	0.026	0.259	0.007	0.016
u2i	0.012	-0.679	-0.008	-0.016	0.025	-0.324	-0.008	-0.019
u2o	-0.116	0.645	-0.075	-0.145	-0.062	0.658	-0.041	-0.099
o2f	-0.733	0.776	-0.569	-1.105	-0.648	1.766	-1.145	-2.776
o2i	0.668	0.578	0.386	0.750	0.548	0.747	0.409	0.991
o2u	-0.822	0.041	-0.034	-0.066	-0.666	-0.017	0.011	0.028
δ	.	.	0.515	.	.	.	0.412	.

Notes: Controls are: post 2019 dummy, linear trend, and their interaction; seasonal quarterly dummies.

worker flows. This suggests that the IPW technique is increasing the weight for individuals whose informality is more persistent and who react less to transitions.

Table 2: Decomposition of informality-population ratio persistence, survey weights

	no DFL adjustment				DFL adjustment			
transition	α	β	$\alpha\beta$	S	α	β	$\alpha\beta$	S
f2i	0.873	0.624	0.545	1.019	0.793	1.081	0.857	2.093
f2u	0.366	0.234	0.086	0.160	0.215	0.380	0.082	0.199
f2o	0.344	1.122	0.386	0.721	0.418	2.245	0.938	2.290
i2f	0.148	0.466	0.069	0.129	0.073	1.319	0.096	0.235
i2u	0.095	0.229	0.022	0.040	-0.070	0.293	-0.020	-0.050
i2o	-0.256	0.505	-0.129	-0.242	-0.297	1.509	-0.448	-1.094
u2f	-0.017	-0.077	0.001	0.002	0.007	0.426	0.003	0.007
u2i	-0.008	-0.590	0.005	0.009	0.006	-0.305	-0.002	-0.004
u2o	-0.141	0.603	-0.085	-0.159	-0.084	0.730	-0.062	-0.150
o2f	-0.806	0.979	-0.789	-1.476	-0.735	2.143	-1.576	-3.847
o2i	0.721	0.662	0.477	0.892	0.611	0.886	0.541	1.321
o2u	-0.948	0.054	-0.051	-0.096	-0.771	-0.000	0.000	0.001
δ	.	.	0.535	.	.	.	0.410	.

Note: Controls are: post 2019 dummy, linear trend, and their interaction; seasonal quarterly dummies.

Table 3 explores the impact of missing the effect of the COVID-19 pandemic on the persistence decomposition. The overall effect is a sizable increase in the measured persistence of informality, from $\delta = 0.515$ to $\delta = 0.594$, a 15% larger coefficient. When DFL adjustments take place, the persistence also decreases with respect to the case with varying population composition, but it is still larger than the persistence found in the comparable exercises in Tables 1 and 2. A priori, not accounting for the COVID-19 shock is expected to have a large influence on the results, as it is a major outlier in the series, as shown in Section 4.1. Nevertheless, the picture is not that different: the indirect outflow $o2f$ is still the most influential transition in absolute terms. Informality inflows from formality, $f2i$ and from inactivity, $o2i$ also account for a large part of informality persistence. Unlike the baseline case in Table 1, the flow from formal jobs to inactivity, $f2o$ becomes more important. The numbers on the right panel showing the results for DFL-weighted flows show very large impacts for both $f2o$ and $o2f$ flows, with opposite signs.

Table 3: Decomposition of informality-population ratio persistence, IPW

	No DFL weights				DFL weights			
transition	α	β	$\alpha\beta$	S	α	β	$\alpha\beta$	S
f2i	0.836	0.488	0.408	0.687	0.520	0.846	0.440	0.892
f2u	0.583	0.258	0.150	0.253	0.410	0.319	0.131	0.265
f2o	0.528	0.875	0.462	0.778	0.943	2.079	1.961	3.978
i2f	0.034	0.313	0.011	0.018	-0.203	1.329	-0.270	-0.548
i2u	0.048	0.132	0.006	0.011	-0.102	0.114	-0.012	-0.024
i2o	-0.339	0.151	-0.051	-0.086	-0.489	1.076	-0.526	-1.067
u2f	0.013	-0.005	-0.000	-0.000	0.029	0.858	0.025	0.051
u2i	0.011	-0.279	-0.003	-0.005	0.003	0.013	0.000	0.000
u2o	-0.154	-0.128	0.020	0.033	-0.134	-0.372	0.050	0.101
o2f	-0.826	0.826	-0.683	-1.150	-0.832	2.297	-1.912	-3.879
o2i	0.634	0.534	0.339	0.571	0.719	0.968	0.697	1.413
o2u	-0.927	0.071	-0.066	-0.111	-0.857	0.105	-0.090	-0.183
δ	.	.	0.594	.	.	.	0.493	.

Note: Controls are seasonal quarterly dummies.

In the appendix A.2 I also report the case of decomposition without acknowledging non-response nor COVID-19 structural break adjustments in Table A3. The results show that both changes with respect to the baseline case I portrayed in Table 1 increase the persistence of the informality.

5 Conclusions

Researchers have mainly focused on understanding the determinants of informality, adopting a static perspective. While the prevalence of informality is undoubtedly important, understanding its dynamics is key. Whether a worker stays for a short or long time in an informal job, often linked to low wages and unpleasant working conditions, has strong implications for welfare and inequality. In this paper, I provide a simple analytical framework to attribute the observed persistence of informality to worker flows. To do so, I rely on a first-order Markov approximation and use an inverse probability weighting (IPW) to correct for non-response in the rotation panel of the employment survey. The latter may be important, as non-response is closely linked to characteristics often observed in informal workers.

I find that the measured average informality persistence seems substantially lower compared to the available evidence for big Latin American economies: Argentina, Brazil, and Mexico (Bosch and Maloney, 2010). Using the decomposition, I can attribute a great deal of the persistence of informality in Chile to inflows from formality ($f2i$) and inactivity ($f2o$), as well as to indirect outflows such as the inactivity-to-formality transition, $o2f$. A failure to account for either the non-response of interviewees and/or the atypical behavior after the pandemic's beginning leads to an overestimation of the informality's persistence. An implication is that policies preventing workers from falling into informality are likely more effective at reducing its average duration than policies pulling workers from informality.

Keeping constant the composition of the Chilean population in 2017 via DFL weighting reduces the informality persistence in all studied cases. This shows that the change in composition that has occurred in Chile has increased persistence, as the population share of groups such as immigrants has substantially increased. As a consequence, if the underlying reasons behind the compositional change, such as immigration policies, population aging, etc., remain, informality persistence may increase in the future.

References

Akay, A. and M. Khamis (2012, Jan). *The Persistence of Informality: Evidence from Panel Data*, Volume 34 of *Research in Labor Economics*, Chapter 7, pp. 229–255. Emerald Group Publishing Limited.

Bennett, M. (2021). All things equal? Heterogeneity in policy effectiveness against COVID-19 spread in Chile. *World Development 137*, 105208.

Bernhardt, R., D. Munro, and E. Wolcott (2021). How Does the Dramatic Rise of CPS Non-Response Impact Labor Market Indicators? Technical report, GLO Discussion Paper.

Bosch, M. and W. F. Maloney (2010). Comparative analysis of labor market dynamics using Markov processes: An application to informality. *Labour Economics 17*(4), 621–631.

Busso, M., J. Camacho, J. Messina, and G. Montenegro (2021). Social protection and informality in Latin America during the COVID-19 pandemic. *PLOS ONE*.

Choi, S., A. Janiak, and B. Villena-Roldán (2015). Unemployment, Participation and Worker Flows Over the Life-Cycle. *The Economic Journal 125*(589), 1705–1733.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica 64*(5), 1001–1044.

Elsby, M. W., B. Hobijn, and A. Şahin (2015). On the importance of the participation margin for labor market fluctuations. *Journal of Monetary Economics 72*, 64–82.

Elsby, M. W. L., R. Michaels, and G. Solon (2009). The Ins and Outs of Cyclical Unemployment. *American Economic Journal: Macroeconomics 1*(1), 84–110.

Escobedo, J. and J. Moreno (2020). Transición y Persistencia en el Ciclo Formal-Informal en México. *Revista de Economía Laboral. 17*(1), 1–45.

Gong, X., A. van Soest, and E. Villagomez (2004). Mobility in the Urban Labor Market: A Panel Data Analysis for Mexico. *Economic Development and Cultural Change 53*(1), 1–36.

Gozzi, N., M. Tizzoni, M. Chinazzi, L. Ferres, A. Vespignani, and N. Perra (2021). Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile. *Nature communications 12*(1), 2429.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public opinion quarterly 70*(5), 646–675.

Günther, I. and A. Launov (2012). Informal employment in developing countries: Opportunity or last resort? *Journal of Development Economics 97*(1), 88–98.

Haanwinckel, D. and R. R. Soares (2021). Workforce composition, productivity, and labour regulations in a compensating differentials theory of informality. *The Review of Economic Studies 88*(6), 2970–3010.

Kucera, D. and T. Xenogiani (2009). *Persisting Informal Employment: What Explains It?* Is Informal Normal ?: Towards More and Better Jobs in Developing Countries.

Maloney, W. F. (1999). Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico. *The World Bank Economic Review 13*(2), 275–302.

Neumark, D. and D. Kawaguchi (2004). Attrition bias in labor economics research using matched CPS files. *Journal of Economic and Social Measurement 29*(4), 445–472.

Shimer, R. (2012, April). Reassessing the Ins and Outs of Unemployment. *Review of Economic Dynamics 15*(2), 127–148.

Shimer, R. (2013). *Job Search, Labor Force Participation, and Wage Rigidities*, Volume 2 of *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress*, Chapter 5, pp. 197–235. Cambridge University Press.

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of econometrics 141*(2), 1281–1301.

A Appendix

A.1 Methodological note on ENE survey data

The identification variables between December 2019 and March 2020 present inconsistencies that were corrected by means of a file that allows linking the identifiers of these surveys, obtained through a formal request to this the *Instituto Nacional de Estadísticas* (INE).

A similar situation occurs between December 2020 and March 2021, although this inconsistency could be corrected by means of an identifier built on the basis of the variables `estrato` (stratum), `id_identificacion`, `parentesco` (kinship), `hogar` (household), `nro_linea` (number of line), `sexo` and `edad` (age). These arrangements worked correctly with Stata format files accessed on the INE website https://www.ine.cl/estadisticas/sociales/mercado-laboral/ocupacion-y-desocupacion in January 2023, combining annual files. For further details or access to codes used, write to benjamin@benjaminvillena.com

A.2 Additional tables

Table A1: Logit models for DFL and IPW adjustments

	(1)	(2)
Dependent variable	Year 2017	Response next quarter
covariates		
female	0.217***	0.103***
	(0.0608)	(0.0356)
Chilean	-0.0396	0.311***
	(0.0392)	(0.0216)
female × Chilean	-0.0910*	-0.0325
	(0.0522)	(0.0293)
head	0.146***	0.153***
	(0.0126)	(0.00797)
female × head	-0.0968***	-0.0378***
	(0.0164)	(0.0108)
married	-0.0339	0.112***
	(0.0293)	(0.0183)
partner not married	0.0348	-0.0267
	(0.0313)	(0.0195)
single	0.0303	0.0748***
	(0.0314)	(0.0194)
divorced	0.0434	-0.0481**
	(0.0348)	(0.0217)
female × married	0.00343	0.0692***
	(0.0331)	(0.0210)
female × partner	-0.0341	0.0662***
	(0.0354)	(0.0223)
female × single	-0.139***	-0.0256
	(0.0342)	(0.0215)
female × divorced	-0.110***	0.0481*
	(0.0390)	(0.0247)
Observations	3,732,932	3,732,932

Note: Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$. The estimated logit models have a dependent variable in column (1) a binary that takes value 1 if the individual is interviewed in the next quarter. In column (2), the dependent variable takes value 1 if the observation took place in 2017.
Other controls not shown in the table are categorical variables for age group (18-29, 30-39, 40-49, 50-59, and 60+); educational level (primary, scientific-humanities high-school, technical high-school, technical tertiary, college, and graduate); geographical zone (north, south, Santiago); and cohort of birth (before 1939, 1940-44, 1945-49,...,1990-94, and after 1995).

Table A2: Descriptive statistics of stock and flow variables by quarter

description	variable	weights	mean	sd	min	max	count
Informal job to Employment ratio	i/e	survey	0.380	0.0214	0.305	0.416	47
Informal job to Employment ratio	i/e	IPW	0.378	0.0212	0.304	0.415	47
Informal job to Employment ratio	i/e	IPW-DFL	0.382	0.0254	0.304	0.434	47
Informal job to Population ratio	i/a	survey	0.227	0.0223	0.146	0.248	47
Informal job to Population ratio	i/a	IPW	0.227	0.0218	0.146	0.247	47
Informal job to Population ratio	i/a	IPW-DFL	0.225	0.0212	0.146	0.245	47
Unemployed to Labor Force ratio	$u/(e+u)$	survey	0.073	0.0146	0.056	0.123	47
Unemployed to Labor Force ratio	$u/(e+u)$	IPW	0.075	0.0151	0.057	0.126	47
Unemployed to Labor Force ratio	$u/(e+u)$	IPW-DFL	0.076	0.0154	0.059	0.127	47
Inactive to Population ratio	o/a	survey	0.356	0.0272	0.336	0.456	47
Inactive to Population ratio	o/a	IPW	0.353	0.0258	0.330	0.450	47
Inactive to Population ratio	o/a	IPW-DFL	0.365	0.0324	0.322	0.450	47
Informal job to Formal job trans rate	$i2f$	survey	0.110	0.0362	0.072	0.343	47
Informal job to Formal job trans rate	$i2f$	IPW	0.112	0.0366	0.075	0.348	47
Informal job to Formal job trans rate	$i2f$	IPW-DFL	0.111	0.0363	0.073	0.346	47
Informal job to Informal job trans rate	$i2i$	survey	0.691	0.0788	0.219	0.742	47
Informal job to Informal job trans rate	$i2i$	IPW	0.689	0.0783	0.219	0.741	47
Informal job to Informal job trans rate	$i2i$	IPW-DFL	0.685	0.0801	0.205	0.741	47
Informal job to Unemployed trans rate	$i2u$	survey	0.039	0.0094	0.023	0.082	47
Informal job to Unemployed trans rate	$i2u$	IPW	0.040	0.0096	0.024	0.084	47
Informal job to Unemployed trans rate	$i2u$	IPW-DFL	0.040	0.0102	0.023	0.084	47
Informal job to Inactive trans rate	$i2o$	survey	0.160	0.0473	0.133	0.389	47
Informal job to Inactive trans rate	$i2o$	IPW	0.159	0.0463	0.132	0.383	47
Informal job to Inactive trans rate	$i2o$	IPW-DFL	0.163	0.0473	0.131	0.393	47
Formal job to Formal job trans rate	$f2f$	survey	0.869	0.0741	0.380	0.906	47
Formal job to Formal job trans rate	$f2f$	IPW	0.869	0.0737	0.383	0.904	47
Formal job to Formal job trans rate	$f2f$	IPW-DFL	0.865	0.0739	0.380	0.907	47
Formal job to Informal job trans rate	$f2i$	survey	0.059	0.0214	0.029	0.192	47
Formal job to Informal job trans rate	$f2i$	IPW	0.059	0.0214	0.030	0.192	47
Formal job to Informal job trans rate	$f2i$	IPW-DFL	0.060	0.0214	0.031	0.192	47
Formal job to Unemployed trans rate	$f2u$	survey	0.026	0.0064	0.020	0.059	47
Formal job to Unemployed trans rate	$f2u$	IPW	0.027	0.0066	0.020	0.061	47
Formal job to Unemployed trans rate	$f2u$	IPW-DFL	0.027	0.0072	0.020	0.064	47
Formal job to Inactive trans rate	$f2o$	survey	0.045	0.0493	0.027	0.369	47
Formal job to Inactive trans rate	$f2o$	IPW	0.045	0.0486	0.027	0.364	47
Formal job to Inactive trans rate	$f2o$	IPW-DFL	0.047	0.0485	0.029	0.364	47
Unemployed to Formal job trans rate	$u2f$	survey	0.217	0.0376	0.102	0.352	47
Unemployed to Formal job trans rate	$u2f$	IPW	0.218	0.0384	0.102	0.358	47
Unemployed to Formal job trans rate	$u2f$	IPW-DFL	0.217	0.0393	0.099	0.369	47
Unemployed to Informal job trans rate	$u2i$	survey	0.234	0.0302	0.123	0.288	47
Unemployed to Informal job trans rate	$u2i$	IPW	0.234	0.0306	0.123	0.291	47
Unemployed to Informal job trans rate	$u2i$	IPW-DFL	0.229	0.0306	0.122	0.287	47
Unemployed to Unemployed trans rate	$u2u$	survey	0.277	0.0462	0.069	0.384	47
Unemployed to Unemployed trans rate	$u2u$	IPW	0.278	0.0466	0.071	0.389	47
Unemployed to Unemployed trans rate	$u2u$	IPW-DFL	0.274	0.0485	0.064	0.383	47
Unemployed to Inactive trans rate	$u2o$	survey	0.273	0.0382	0.218	0.437	47
Unemployed to Inactive trans rate	$u2o$	IPW	0.271	0.0379	0.215	0.434	47
Unemployed to Inactive trans rate	$u2o$	IPW-DFL	0.279	0.0379	0.220	0.436	47
Inactive to Formal job trans rate	$o2f$	survey	0.042	0.0488	0.022	0.367	47
Inactive to Formal job trans rate	$o2f$	IPW	0.043	0.0497	0.023	0.375	47
Inactive to Formal job trans rate	$o2f$	IPW-DFL	0.042	0.0499	0.022	0.375	47
Inactive to Informal job trans rate	$o2i$	survey	0.099	0.0172	0.048	0.189	47
Inactive to Informal job trans rate	$o2i$	IPW	0.099	0.0170	0.048	0.188	47
Inactive to Informal job trans rate	$o2i$	IPW-DFL	0.098	0.0164	0.048	0.186	47
Inactive to Unemployed trans rate	$o2u$	survey	0.037	0.0047	0.029	0.050	47
Inactive to Unemployed trans rate	$o2u$	IPW	0.038	0.0050	0.029	0.054	47
Inactive to Unemployed trans rate	$o2u$	IPW-DFL	0.038	0.0056	0.028	0.056	47
Inactive to Inactive trans rate	$o2o$	survey	0.823	0.0652	0.396	0.900	47
Inactive to Inactive trans rate	$o2o$	IPW	0.820	0.0658	0.389	0.898	47
Inactive to Inactive trans rate	$o2o$	IPW-DFL	0.822	0.0660	0.391	0.896	47

Table A3: Decomposition of informality-population ratio persistence, survey weights

	No DFL weights				DFL weights			
transition	α	β	$\alpha\beta$	S	α	β	$\alpha\beta$	S
f2i	0.858	0.437	0.375	0.602	0.522	0.794	0.414	0.825
f2u	0.584	0.250	0.146	0.235	0.424	0.293	0.124	0.247
f2o	0.564	0.769	0.434	0.697	1.003	2.000	2.006	3.990
i2f	0.033	0.220	0.007	0.012	-0.209	1.253	-0.262	-0.521
i2u	0.014	0.147	0.002	0.003	-0.116	0.099	-0.011	-0.023
i2o	-0.318	0.056	-0.018	-0.028	-0.473	0.980	-0.463	-0.921
u2f	0.010	0.025	0.000	0.000	0.039	0.896	0.035	0.070
u2i	-0.002	-0.261	0.000	0.001	-0.011	0.013	-0.000	-0.000
u2o	-0.169	-0.254	0.043	0.069	-0.148	-0.426	0.063	0.125
o2f	-0.913	0.727	-0.663	-1.066	-0.945	2.201	-2.080	-4.139
o2i	0.707	0.530	0.375	0.603	0.807	0.980	0.790	1.572
o2u	-1.007	0.079	-0.079	-0.128	-0.902	0.125	-0.113	-0.224
δ	.	.	0.622	.	.	.	0.503	.

Note: Controls are seasonal quarterly dummies.