# Unpacking the persistence of informality*

Benjamín Villena-Roldán

Universidad Andres Bello, Institute for Economic Policies;

$LM^2C^2$; and MIPP

April 23, 2024

**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 is important for developing effective public policies. Using a Markov representation of worker flows and correcting for time aggregation, I find low persistence of informality in Chile, implying an average duration of nearly 3.5 months, and estimate the contributions of labor market transitions among formality, informality, unemployment, and inactivity. The flow into informality from unemployment is the main force accounting for persistence, which suggests that informality is a temporary shelter from joblessness. I also find informality persistence is higher for females, young workers, and tertiary-educated individuals.

*Keywords:* Informality, unemployment, transition rates, time-aggregation bias..

*JEL Codes:* J46, J64, E24

*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.

# 1 Introduction

In developing economies, a large proportion of the workforce is informal, which typically implies a high share of jobs without social security protection and often poor career prospects. Informal jobs are generally provided by low-productivity firms that evade tax and social security payments. Nevertheless, on top of these relatively traditional concerns, the COVID-19 pandemic showed that governments, especially in developing countries, had a hard time reaching informal workers with subsidies or in-kind transfers to reduce their mobility and assuage the disease's spread. The media[1] and academic research show that the COVID-19 lockdowns made it difficult for informal workers to get government transfers (Busso et al., 2021).

While the level of informal employment is obviously important, a key issue is whether informality is permanent or transitory. If informality is a way to avoid unemployment, generates low living standards, and shows high persistence, it becomes a trap for workers that appropriate policies could mitigate. On the other hand, if informality provides a temporary shelter from unemployment (Fernández and Meza, 2015; Donovan et al., 2023) but workers can transition to formal jobs easily, the informal sector may provide some self-insurance, especially in economies with poor social security or restrictive credit access.

The measurement of informality persistence depends on the level of aggregation of the data and substantially differs across different groups of the population. There are two main approaches in the literature to empirically assess labor market transitions. Both of them have pros and cons. On one hand, the macro tradition, using a Markov chain approach to understand labor market dynamics, uses aggregated data to correct for time-aggregation bias (Perry, 1972; Shimer, 2012; Elsby et al., 2009; Gomes, 2015), misclassification (Poterba and Summers, 1986; Feng and Hu, 2013; Elsby et al., 2015), or both (Choi et al., 2015; Elsby et al., 2015) often with no or little heterogeneity considerations. On the other hand, there are microeconometric studies that focus on individual-level transitions and provide estimates highlighting the heterogeneous responses across demographic groups in Mexico (Gong et al., 2004; Escobedo and Moreno, 2020), Ukraine (Akay and Khamis, 2012), Côte

[1] See, for instance, BBC Latin America

d'Ivoire (Günther and Launov, 2012),and four Sub-Saharian countries (Danquah et al., 2021). However, these micro-level studies cannot correct for time-aggregation bias.

In this paper, I propose to bridge both views by constructing a synthetic panel (Verbeek, 2008) of transition probabilities, previously corrected for time-aggregation bias, per bins defined by gender, age, and education. Taking time-aggregation into account is important because actual transitions in labor markets may occur at a higher frequency and cannot be observed since the survey data is elicited at a lower frequency in quarterly rotating panels, as in Chile and other countries. Bosch and Maloney (2010); Gomes (2015) apply these techniques to informality worker transitions to better understand the dynamics of labor markets.

I apply this method to the *Encuesta Nacional de Empleo* (ENE henceforth), a Chilean National Employment Survey covering 2010–2023, which has a rotating panel structure allowing the computation of quarterly transitions. I also use the weighting technique of DiNardo et al. (1996) (DFL henceforth) to construct transitions as if the composition were that of 2017, the year of the unemployment rate closest to the pre-pandemic average. I show a decomposition of the persistence, i.e., the marginal change of informality in the next period when informality increases in the present, into different inflows to informality from formal employment, $f$, unemployment, $u$, and inactivity, $o$. Using weekly, time-bias-corrected probability transitions, I estimate a coefficient implying an average duration of informality spells of 15 weeks, or approximately 3.5 months. The decomposition shows that the unemployment-to-informality ($ui$) and the inactivity-to-informality-to ($oi$) flows significantly account for 45% and 20% of the persistence of informality, respectively. Before the "Social Unrest" (October 2019) and the COVID pandemic, these numbers reached 57% and 22%, respectively.

In contrast, if time aggregation bias is not corrected, I estimate an average duration of informality of 2.5 quarters, i.e., 7.5 months. The persistence is still mainly accounted for by the $ui$ and $oi$ transitions, but the share attributed to increased informality-staying flows, $ii$, increases and becomes significant. The estimated persistence in Chile implies, in any case, substantially shorter informality durations than those estimated for other emergent and developing economies. Bosch and Maloney (2010) estimate for Argentina, Mexico, and Brazil an average duration of one year for a salaried informal job and two years for self-employment. Using Brazilian data, Haanwinckel and Soares (2021) obtain an implied average duration for unskilled informal workers of around 13 months. For Colombia, Mondragón-Vélez et al. (2010) report that 80% of informal workers remain informal after a year. Compared to Latin American findings, African economies (South Africa, Tanzania, Ghana, and Uganda) show much lower persistence in formal jobs and a very high persistence in formal and informal self-employment (Danquah et al., 2021). For Egypt, Tansel and Ozdemir (2019) find substantial persistence in informality during a six-year period. However, these estimates are hardly directly comparable due to the development gap between African and Latin American economies and the different frequency of data collected: quarterly in Latin American countries and ranging between two and four years for African economies. For a sample of 49 countries, Donovan et al. (2023) show that employment exit rates are much higher for informal jobs and that the informal-to-formal gap widens as per capita income declines. Hence, lower informality rates and shorter informality durations in Chile are expected, as the latter authors suggest, because the Chilean economy has one of the highest per capita income levels in Latin America.

I also compute how heterogeneity varies persistence across gender, age, and educational groups. There is an increasing persistence of informality over the years, reaching a peak in 2020, and a subsequent reversal to average levels during 2021-23. Females, young people, and individuals with tertiary education show greater persistence compared to their corresponding counterparts. The fact that more educated workers exhibit high persistence suggests that their jobs are likely of high quality, despite being informal. Finally, I show that the results are robust to several dynamic specifications, including transition probability lags and different strategies to control for the COVID pandemic disruption. Moreover, I also show that the results barely depend on correcting for composition via DFL weighting or just using the original sample weights.

I also report labor market stocks and the weekly transition series for 2010–2023 using survey and DFL weights. 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 probabilities 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 abrupt spikes and falls in transitions for all labor market states.

All in all, the findings suggest that unemployed people in Chile actively look for informal work opportunities that are more often temporary. The data shows a lively movement between informal and formal jobs. Even if the underlying assumptions behind the time-aggregation weekly correction do not hold, the unadjusted data still shows that informality is clearly less persistent in Chile than it is reported for other emergent economies. Informal work does not last very long, especially for unskilled workers. So, policies aimed at reducing informal work might need to be careful, because for many workers, informality is a temporary way to secure some income while they look for better job opportunities.

The rest of the paper is as follows: In Section 2, I discuss the informality definition used in this paper, describe salient features of the data, and show how to construct series while taking their compositional changes into account. In Section 3, I explain the Markov chain setup and how to correct time-aggregation bias. I also decompose stock persistence into worker flow contributions and show how some key quantities can be empirically estimated using local projections. Section 4 shows the main results. First, I plot and explain stock (shares) and flow variables using survey and DFL weights. Second, I show the average persistence decomposition of informality, report the heterogeneous persistence across demographic groups, and discuss its interpretation. Third, I discuss some robustness checks. Finally, Section 5 presents the conclusions.

## 2 Data

In this paper, I use the *Encuesta Nacional de Empleo* (ENE), administered by the *Instituto Nacional de Estadísticas* (INE), to compute the official statistics of the labor market in Chile. [2] The 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. The sampling design of ENE contains data on surveyed individuals within the last moving quarter to obtain a smoother report and greater temporal accuracy. In 2010, the survey was reformulated to meet the standards of the Organization for Economic

[2] All the data used in this paper is available at https://www.ine.cl/estadisticas/sociales/mercado-laboral/ocupacion-y-desocupacion

Co-operation and Development (OECD).

### 2.1 Definitions

Following the definition provided by INE, the *employed* are individuals who are older than 18 and worked at least one hour in exchange for wages or benefits during the week before the interview. INE's official definition of informality implements guidelines from the International Labor Organization (ILO). Unfortunately, some of the data used was collected only since 2017, especially information regarding the tax status of the employer. In consequence, I construct a measurement that can be computed from March 2010 onward and is simpler.

The definition I adopt here also stresses a worker perspective and highlights the lack of legal regulation coverage (Bosch and Maloney, 2010; Akay and Khamis, 2012; Günther and Launov, 2012). In consequence, informal workers lack legal work contracts, *and* their employers fail to contribute to the pension and health system (PHS), which is mandated by Chilean law. Employers (business owners) and self-employed are considered formal as long as they pay PHS contributions[3]; if not, they are also formal if they work at a likely formal location (plant, store, etc.) *and* work for a firm with at least five employees. Unpaid family workers are never considered formal. Naturally, *informal* workers are non-formal employed individuals. For the sake of completeness, I state additional definitions, too: The *unemployed* are persons of working age who were not employed during the reference week, had carried out job search activities during the last four weeks, and are available for work within the next two weeks. *Inactive* individuals are those older than 18 who are neither employed nor unemployed.

There are a myriad of informality definitions with statistical or conceptual purposes, but no one is a panacea (Dell'Anno, 2022). The definition I adopt here is pragmatic since it can be implemented with the available data and aligns with ILO definitions, which is similar to the one that has been used in other studies (Günther and Launov, 2012; Bosch and Esteban-Pretel, 2012; Fernández and Meza, 2015; Danquah et al., 2021). It is also a worker-centered perspective in contrast with a firm-centered approach to an informality definition, which considers persons working for firms not complying with legal regulations as informal workers

[3]The same applies to home service workers

(Perry et al., 2007; Ulyssea, 2020). In practice, informal firms are often identified as those with few employees (Maloney, 1999; Gong et al., 2004). Hence, the worker focus refers to key aspects of informality linked to the worker's welfare: access to social security, health, and legal protection. Moreover, as informality data comes from surveys whose respondents are workers, they are more likely aware of their employer's compliance with social security and health payments that directly affect them than their employer's corporative tax status. For these reasons, the worker-declared information regarding PHS or the work contract seems more reliable. In any case, often different measurements of informality show a similar picture. For instance, Mondragón-Vélez et al. (2010) for Colombia and Fernández and Meza (2015) for Mexico show that several informality definitions are highly correlated to one another and display similar cyclical behavior.

### 2.2 Descriptive statistics

Table 1 contains descriptive statistics from the National Employment Survey for individuals aged 25–65, comparing statistics before March 2020 and after March 2020, when the COVID pandemic started in Chile. The table considers four previously defined employment statuses and calculates the averages and relative frequencies of different variables of interest for both time periods.

Men accounted for nearly 60% of formal employment before the pandemic, which declined somewhat in the post-COVID period. Both the unemployed and the inactive had higher male participation after the pandemic. All employment shows a dramatic increase in the share of foreign workers, which is consistent with a huge increase in the immigration flow to Chile from 2018 onwards, particularly from Venezuela, Colombia, and Haiti. The share of foreign workers in formal employment more than doubles after 2020, and informal employment more than triples. The share of unemployed and inactive foreigners roughly doubles as well. The percentage of heads of household notably increases for formal and informal workers in the post-COVID period, which contrasts with the change shown by the unemployed and the inactive.

Before and after the pandemic, informal workers had a higher average age than formal workers. Only the inactive surpassed the average age of informal workers. After the pandemic's inception, the distribution across age groups remained roughly the same. In contrast, informal worker shares by age group reveal an increase for the youngest groups (25–38). An opposite phenomenon is observed for the unemployed, whose participation increases after COVID. Finally, inactives concentrate more on middle-aged workers (39–51) in comparison to other groups.

The distribution of education groups exhibits notable changes after COVID. The share of workers with less than high school declines for all employment categories except for inactivity. The same holds for the population as a whole (21.3% vs. 15.7%). This reflects a sustained increase in educational attainment for younger cohorts and immigrants compared to that of retirees during the 2010s. This share decline is particularly acute for informal workers, with almost 10 percentage points. In turn, the share of regular high school achievers mildly decreases in formal employment and unemployment after COVID but increases in informal jobs. The technical high school category showed a slight decrease across all labor states, including their share in the population. Technical tertiary education increases their overall participation by 1.5 percentage points, with increments in both formal and informal jobs. College-educated persons exhibit a remarkable overall increase (3.8 percentage points) and a notorious increase of 4.9 percentage points of their share in formal employment and 5.4 percentage points increase in informal employment. Graduate-educated workers, while scarcer in the economy, have witnessed a large increase in their share as well.

The hours worked by informal workers are substantially lower than those of formal workers, with hours worked dropping after the pandemic to 1.3 weekly hours on average. The reported tenure among formal workers before COVID reached 86 months on average (7.2 years), which contrasts with that of informal workers (5.7 years). After COVID, both numbers substantially declined, especially among the informal. These facts are consistent with a large increase in transitions to inactivity among the elderly, a large increase in immigration, and a large spike in job destruction at the onset of the onset of the pandemic.. The share of private salaried employment slightly increased after the pandemic, and the percentage of informal employment declined. Public employment, almost always formal, increased its share by 2.4 percentage points overall after March 2020. The statistics show that self-employment is mostly informal, and that pattern was exacerbated in the post-COVID period. Employers as well as other types of jobs, such as domestic service workers and unpaid family workers, also exhibit notable declines.

Table 1: National Employment Survey (ENE): Descriptive Statistics

| | pre COVID (March 2010 - February 2020) labor status | | | | | post COVID (March 2020 - December 2023) labor status | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Formal | Informal | Unempl. | Inactive | Total | Formal | Informal | Unempl. | Inactive | Total |
| *Demographics* | | | | | | | | | | |
| male | 59.9 | 53.1 | 52.8 | 20.6 | 49.5 | 57.9 | 55.2 | 56.5 | 27.4 | 49.8 |
| foreign | 5.1 | 4.7 | 5.1 | 3.0 | 4.5 | 10.9 | 13.6 | 10.4 | 5.9 | 10.0 |
| head | 53.9 | 45.7 | 29.3 | 16.8 | 43.0 | 59.8 | 51.3 | 34.0 | 23.5 | 47.8 |
| *Age* | | | | | | | | | | |
| age (avg) | 41.8 | 43.9 | 38.5 | 44.9 | 42.7 | 41.9 | 43.3 | 39.7 | 45.1 | 42.8 |
| age 25-38 (%) | 42.7 | 35.3 | 56.4 | 34.9 | 40.3 | 42.9 | 38.7 | 52.0 | 35.3 | 40.9 |
| age 39-51 (%) | 35.1 | 35.3 | 28.1 | 29.1 | 33.4 | 34.2 | 33.6 | 29.7 | 26.8 | 32.0 |
| age 52-65 (%) | 22.2 | 29.4 | 15.5 | 36.0 | 26.3 | 22.8 | 27.6 | 18.3 | 37.8 | 27.0 |
| *Education* | | | | | | | | | | |
| less than HS (%) | 15.0 | 29.8 | 15.2 | 31.9 | 21.3 | 9.9 | 20.1 | 11.9 | 26.4 | 15.7 |
| regular HS (%) | 29.3 | 36.4 | 32.4 | 33.3 | 31.5 | 27.9 | 38.1 | 33.5 | 36.3 | 31.9 |
| tech. HS (%) | 13.2 | 12.1 | 11.7 | 10.5 | 12.4 | 10.9 | 11.6 | 10.4 | 10.2 | 10.8 |
| tech. tertiary (%) | 14.3 | 9.2 | 14.4 | 9.4 | 12.3 | 15.9 | 11.0 | 15.0 | 10.7 | 13.8 |
| college (%) | 25.1 | 11.7 | 24.8 | 14.1 | 20.4 | 30.0 | 17.1 | 27.0 | 15.5 | 24.2 |
| graduate (%) | 3.1 | 0.8 | 1.5 | 0.7 | 2.1 | 5.4 | 2.2 | 2.2 | 0.9 | 3.6 |
| *Work variables* | | | | | | | | | | |
| week hours (avg) | 44.9 | 34.8 | - | - | 42.6 | 43.6 | 33.1 | - | - | 41.3 |
| tenure ms (avg) | 86.2 | 68.5 | - | - | 82.1 | 58.0 | 32.0 | - | - | 52.1 |
| private salaried (%) | 67.3 | 31.4 | - | - | 43.0 | 69.9 | 26.0 | - | - | 41.3 |
| public salaried (%) | 15.5 | 0.9 | - | - | 8.8 | 17.9 | 0.9 | - | - | 9.7 |
| self-employed (%) | 9.8 | 50.4 | - | - | 14.0 | 7.6 | 61.1 | - | - | 13.5 |
| employer (%) | 4.3 | 4.3 | - | - | 3.2 | 2.8 | 4.3 | - | - | 2.2 |
| other work (%) | 3.1 | 13.0 | - | - | 3.9 | 1.9 | 7.7 | - | - | 2.2 |
| Observations | 1,199,417 | 368,482 | 92,297 | 552,661 | 2,212,857 | 357,666 | 107,119 | 37,767 | 211,344 | 713,896 |

Note: all the data are freely available at https://www.ine.gob.cl/estadisticas/sociales/mercado-laboral/ocupacion-y-desocupacion. Regular HS (high school) provides formation in science and humanities to prepare students for undergraduate education. Technical HS provides education intended for the labor market or technical tertiary education. Mean hours worked are weekly. Tenure are measured in months.

Table 2 shows another angle of the data that is useful to understand the true nature of informality in Chile. Each row reports the share of workers by educational and formal status that belong to each of the nine big occupational groups in the ISCO08 taxonomy. Groups 1 and 2 are regarded as the most skilled workers in the economy, as they usually complete college or even graduate degrees. Naturally, a large majority of these workers, especially group 2, which covers professional and scientific occupations, hold college diplomas. Most of these workers have formal jobs. Although those without a college degree are less often in formal jobs than their college-educated counterparts, a majority of workers are formal. Moving to the right side of the table, we gradually see groups with presumably less skill. However, even for agricultural workers (group 6) or elementary occupations (group 9), formal jobs are more prevalent. What is more, this is even true for the most unskilled occupations for non-college workers. These show that the Chilean economy is predominantly formal, even in occupations and educational groups in which development economies typically show informality predominance.

Table 2: National Employment Survey (ENE): Shares of occupational groups by education and formality

| | Occupational Group (ISCO08) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Total |
| non-college | | | | | | | | | | |
| Formal (%) | 84.9 | 74.9 | 89.7 | 94.8 | 67.1 | 79.2 | 61.1 | 72.0 | 68.0 | 70.8 |
| Informal (%) | 15.1 | 25.1 | 10.3 | 5.2 | 32.9 | 20.8 | 38.9 | 28.0 | 32.0 | 29.2 |
| Observ. | 12,857 | 7,304 | 70,399 | 82,923 | 261,528 | 80,715 | 228,050 | 160,730 | 433,954 | 1,338,460 |
| college+ | | | | | | | | | | |
| Formal (%) | 94.0 | 91.4 | 90.4 | 95.0 | 74.9 | 83.7 | 69.4 | 61.5 | 68.5 | 86.3 |
| Informal (%) | 6.0 | 8.6 | 9.6 | 5.0 | 25.1 | 16.3 | 30.6 | 38.5 | 31.5 | 13.7 |
| Observ. | 44,134 | 232,728 | 152,373 | 63,113 | 78,569 | 6,253 | 45,279 | 24,074 | 32,517 | 679,040 |
| all | | | | | | | | | | |
| Formal (%) | 92.4 | 91.0 | 90.2 | 94.9 | 69.3 | 79.6 | 62.7 | 70.4 | 68.0 | 77.0 |
| Informal (%) | 7.6 | 9.0 | 9.8 | 5.1 | 30.7 | 20.4 | 37.3 | 29.6 | 32.0 | 23.0 |
| Observ. | 56,991 | 240,032 | 222,772 | 146,036 | 340,097 | 86,968 | 273,329 | 184,804 | 466,471 | 2,017,500 |

Note: all the data are freely available at https://www.ine.gob.cl/estadisticas/sociales/mercado-laboral/ocupacion-y-desocupacion. The occupational groups in the ISCO08 taxonomy are: (1) managers; (2) professionals; (3) technicians and associate professionals; (4) clerical support workers; (5) service and sales workers; (6) skilled agricultural, forestry, and fishery workers; (7) craft and related trades workers; (8) plant and machine operators and assemblers; and (9) elementary occupations.

### 2.3 Transition probabilities

Formally, the ENE survey has probability weights from 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 $x_t$. Using these weights, the number of individuals in state $x$ at time $t$ 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 *survey* estimates in the main paper. 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

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

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 the sample over two periods.

The same definition can be applied to specific groups $g = 1, 2, ..., G$ of the population, rendering the following definition:

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

were $h_{n,t,g}$ is a binary variable that takes value 1 if the individual $n$ at time $t$ belongs to the group $g$ and 0 otherwise.

### 2.4 Composition issues

One concern with the proper estimation of informality persistence is that there have been substantial changes in the composition of the labor force over the time analyzed in Chile. 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 A5 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.

## 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 $xz$. Equation (1) describes the evolution of stocks.

$$\underbrace{\begin{pmatrix} F_t \\ I_t \\ U_t \\ O_t \end{pmatrix}}_{\mathbf{Y}_t} = \underbrace{\begin{pmatrix} 1 - if_t - uf_t - of_t & if_t & uf_t & of_t \\ fi_t & 1 - fi_t - ui_t - oi_t & ui_t & oi_t \\ fu_t & iu_t & 1 - fu_t - iu_t - ou_t & ou_t \\ fo_t & io_t & uo_t & 1 - fo_t - io_t - uo_t \end{pmatrix}}_{\mathbf{Q}_t} \underbrace{\begin{pmatrix} F_{t-1} \\ I_{t-1} \\ U_{t-1} \\ O_{t-1} \end{pmatrix}}_{\mathbf{Y}_{t-1}} \tag{1}$$

which can be compactly expressed as $\mathbf{Y}_t = \mathbf{Q}_t \mathbf{Y}_{t-1}$. Since we focus on the persistence of informality, I derive the following equation for $I_t$

$$I_t = fi_t(F_{t-1} - I_{t-1}) + ui_t(U_{t-1} - I_{t-1}) + oi_t(O_{t-1} - I_{t-1}) + I_{t-1} \tag{2}$$

The last equation highlights the temporal restriction implied by the Markov chain: stocks do not affect each other contemporaneously. This is likely to hold when each period is sufficiently short, like a week. For a substantially longer period, such as a quarter, transitions taking place within such a period generate a mutual feedback effect between different stocks. While equation 2 only directly shows informality inflows ($fi$, $ui$, and $oi$), informality outflows also play a role in informality persistence through the change of the stock of informality in the previous period.

$$
\begin{aligned}
\frac{\partial I_t}{\partial I_{t-1}} - 1 = \rho - 1 = & \underbrace{-I_{t-1}\left(\frac{\partial(fi_t + ui_t + oi_t)}{\partial I_{t-1}}\right)}_{I_{t-1}\frac{\partial ii_t}{\partial I_{t-1}}} + \underbrace{F_{t-1}\frac{\partial fi_t}{\partial I_{t-1}} - fi_t}_{fi \text{ contrib}} \\
& + \underbrace{U_{t-1}\frac{\partial ui_t}{\partial I_{t-1}} - ui_t}_{ui \text{ contrib}} + \underbrace{O_{t-1}\frac{\partial oi_t}{\partial I_{t-1}} - oi_t}_{oi \text{ contrib}} \qquad (3)
\end{aligned}
$$

Equation (3) shows different components in which the persistence quantity, $\rho - 1$, can be decomposed, which is, by no means, unique, but it offers a relatively straightforward accounting framework to think about different factors that affect informality persistence. A first term is just given by the flow-variation of informality staying, e.g., the transition from informality to informality, $ii$, amplified by the share of informal workers. This is just the share of informal workers remaining informal in the next period due to a previous increase in informality. The second term is the contribution of formal workers to informality, which in turn is decomposed into (i) a flow-variation of formality, e.g., the share of formal workers transiting to informality associated with an increase in informality share in the previous period; and (ii) a direct transition effect, the flow $fi$, which reflects the share of workers moving from formal to informal jobs. The third and fourth terms have an analog interpretation as the total contribution to the unemployment-to-informality and inactivity-to-informality transitions ($ui$ and $oi$, respectively).

Equation (3) directly highlights the contributions to persistence of inflows to informality $fi$, $ui$, and $oi$, but this does not mean that the informality outflows are missing. Their importance is captured indirectly by the impact of formality, unemployment, and inactivity shares ($F$, $U$, and $O$, respectively) on the informality share $I$. Consider, for instance, the analog of equation (2) for the dynamic evolution of formality

$$F_t = if_t(I_{t-1} - F_{t-1}) + uf_t(U_{t-1} - F_{t-1}) + of_t(O_{t-1} - F_{t-1}) + F_{t-1}$$

which directly shows how the informality-to-formality $if_t$ flow affects the current share of formality, $F_t$. This is also true for the $uf_t$ and $of_t$ transition probabilities. Hence, the fact that $F_t$ shows up in (3) not only captures the contribution to the $if$ flow into the share of informality but also the contributions of all the other formality inflows. The same argument applies to other labor state inflows $U$ and $O$. Through $U_t$, all unemployment inflows $fu$, $iu$, and $ou$ affect informality in equation (3). The same applies to inactivity inflows $fo$, $io$, and $uo$, which affect $I_t$ by means of $O_t$.

### 3.1 Time aggregation bias

To compute decomposition (3), I need a measurement of transition rates, which can be the quarterly frequency directly obtained from the ENE survey. Nevertheless, a well-known issue with labor market transitions is that they occur at a higher frequency than employment surveys can measure. For instance, the calculation of transition probabilities over a whole quarter misses transitions that occur within shorter time intervals, masking the real underlying dynamics of the labor market. A common way to deal with this problem is using a continuous-time approximation (Shimer, 2012; Bosch and Maloney, 2010); nevertheless, the discrete-time approximation to weekly data (Kaitz, 1970; Perry, 1972) renders a more realistic description of the underlying stochastic process governing labor market dynamics and matches the weekly timing that the survey questions propose to interviewees. As in CPS data and typical employment surveys, the ENE survey also asks the respondent's labor status in the "previous week" so that, for instance, an employment-unemployment-employment transition within a single week should not be captured through this kind of question while a continuous-time approximation assumes it is actually measured (Elsby et al., 2009). For these reasons, I choose a discrete-time approximation using the spectral decomposition theorem as follows:

$$\mathbf{Q}_t = H_t V_t H_t^{-1}$$

where $H$ is the matrix of eigenvectors of $P$ and $\mathbf{Q}_t$ is the diagonal matrix containing the corresponding eigenvalues of $P$. Therefore, assuming constant within-quarter transition rates and considering that each quarter has 13 weeks, the weekly transition matrix is

$$\mathbf{W}_t = H_t V_t^{1/13} H^{-1} \tag{4}$$

If the matrix $\mathbf{Q}_t$ has some zero values for a particular partition of data, I substitute them by a small value $\epsilon = 0.001$ and re-scale each row so that they add up to 1. The purpose of doing this is to reduce the appearance of negative transition probabilities when computing weekly transitions using (4).

### 3.2 Measuring transition-stock derivatives

To compute decomposition (3), I also require to measure the derivatives $\frac{\partial fi_t}{\partial I_{t-1}}$, $\frac{\partial ui_t}{\partial I_{t-1}}$, and $\frac{\partial oi_t}{I_{t-1}}$, which should be linked to some theory regarding the formation of inflows into informality. While it is possible to conjecture an informal job creation process akin to the one in the formal sector, along the lines of canonical models such as in Pissarides (2000) or Shimer (2005), "vacancy-posting" in the informal sector is unobserved. Instead of offering some theory about transitions to and from unemployment and inactivity, I propose a fully empirical approach.

Two empirical polar strategies are generally possible. The first option consists of using individual-level data from the rotating panel structure of the ENE, which yields quarterly transitions, but time-aggregation correction is not possible at the individual level. The second choice is relying on aggregate transition rates, but that misses potentially important cross-sectional heterogeneity. A middle ground between the two ideas is to create a synthetic panel using monthly data reported by the ENE, making it feasible to take care of both aspects altogether (see Verbeek, 2008, for a survey of the literature on pseudo or synthetic panel data).

I thus construct transition rates for different bins in the data defined by age group ($a$), college status ($c$), and gender ($g$). I measure the average partial derivatives using local projections (Jordà, 2005) in the context of a synthetic panel, with fixed effects for each characteristic defining the bin, as follows:

$$yi_{acg,t+1} = \mu_a + \mu_c + \mu_g \; + \alpha I_{acgt} + \sum_x \sum_{z \neq x} \beta^{xz} xz_{acgt} + \delta S_t + \epsilon_{acgt} \quad \forall y \in \{f, u, o\} \tag{5}$$

where the coefficient $\alpha$ measures the empirical estimate of the partial derivative of the flow $yi_{acg,t+1}$ for $y \in \{f, u, o\}$ with respect to the stock of informality[4] at time $t$, $I_{acgt}$ for the bin defined by the categories of age $a$, college status $c$, and gender $g$.

Besides including bin variables fixed effects, $\mu_a$, $\mu_c$, and $\mu_g$ to account for heterogeneity across groups, I also include other time controls in $S_t$ monthly seasonal binary variables. I also explore several specifications in which I vary (i) whether or not to include transitions in $t$ to account for dynamics (restricting $\beta^{xz} = 0$ or not), and (ii) the way to control for the impact of the COVID pandemic in labor market transitions by changing other controls in $S_t$. With respect to the last point, I consider three ways to account for this: (a) a binary variable from March 2020 onwards; (b) a binary variable covering between March 2020 and December 2020, the time period with no vaccines available; and (c) the variable defined in (a) and its interaction with a linear trend. In order to compute the standard error appropriately, I estimate the parameters of the equations defined in (5) using a Seemingly Unrelated Regression (SUR) method (Zellner, 1962) to increase estimation efficiency since the right-hand side variables in the system of equations 5 are not identical in all of them (if they were, equation-by-equation OLS would be optimal).

An important remark is that the obtained coefficients of equations (5) *are not* considered causal estimates but just a method to measure the expected change in transition probabilities when the stock of informal workers changes in the previous quarter. The exercise is in the spirit of other decomposition techniques in labor economics (Fortin et al., 2011) or in macroeconomics (Shimer, 2012; Elsby et al., 2009; Choi et al., 2015).

# 4 Results

## 4.1 Evolution of stocks and flows

Figure 1 depicts series for the adult population using survey and DFL weights, which look very similar. The informal-to-employed ratio declines until mid-2013 and keeps hovering around 25% until the sharp drop of the COVID-19 pandemic. After hitting a bottom near 20%, informality bounces back, reaching 26% by the end of 2021 and remaining at a similar

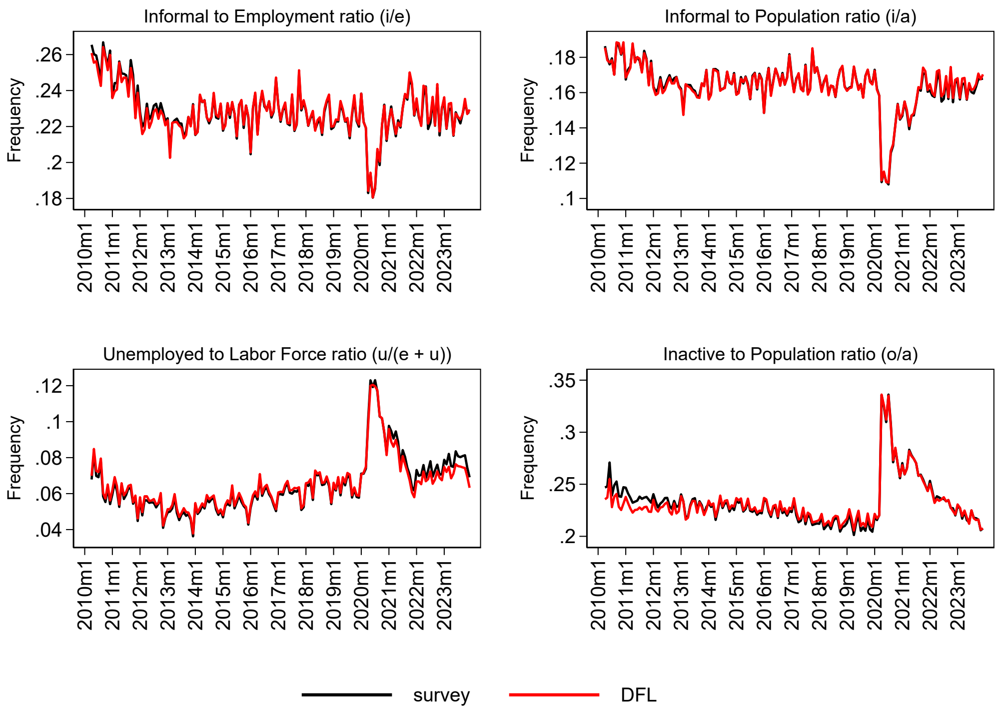
[4] Given the standard normalization of the stocks by the amount of population, $I_{acgt}$ is also the informality-to-population ratio for the bin defined by the group age $a$, college status $c$, and gender $g$.

level until the end of 2023. The informal-to-population ratio ($i/a$) exhibits a qualitatively similar pattern, starting at 17% in 2010 and reaching nearly 14% by the end of 2023. 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% to start increasing again by the end of 2023. The lower-right panel shows a steadily declining inactivity-to-population ratio ($o/a$), until a large increase in March 2020, which slowly and partially reverts to the pre-pandemic level. The observed patterns are common to 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 soon after the first cases, based on the number of contagions reported (Bennett, 2021). These events have clearly disrupted the labor market's functioning since March 2020, generating sudden changes in stocks and flows.

In Figure 1, the DFL-weighted line reveals that the composition of the population apparently has little impact on these aggregate stocks in the economy in general. Nevertheless, the inactive-to-population ratio is visible higher than before 2014 when weighted by DFL, suggesting that part of the decline of this variable during the first years of the sample is due to a compositional change in the population.

Figure 2 depicts informality inflows at the computed weekly frequency. For the $fi$, $ii$, $ui$, and $oi$ flows, there is a quick return to pre-COVID values after 2020. The formality-to-informality probability exhibits a value fluctuating around 0.6% per week and a sudden ten-fold increase at the pandemic onset. The informal-to-informal weekly transition probability also shows a large drop of 15-20 percentage points to rapidly recover to the pre-pandemic level of around 95%. This crude measure suggests an average duration of informality of about $1/(1-0.95) \approx 20$ weeks or 4.7 months. The inflow from unemployment, $ui$, exhibits a large increase from 3% to almost 9-15%, depending on DFL adjustment. Transitions from inactivity to inactivity, $io$, remain hovering around 1% per week until a four-fold increase at the pandemic onset. At that point, the time-aggregation correction renders negative numbers, which reflects the inability of the Markovian framework with time-invariant probability transitions to generate the large drop in the staying probabilities as well as those of the other inflows. As Bosch and Maloney (2010) notice, there is no guarantee that the

Figure 1: Evolution of informality, unemployment, and inactivity ratios




time-aggregation correction provides probabilities strictly in the range between 0 and 1. In the appendix in Figures A13–A16, I provide alternative series reporting moving averages of two and three months that address and solve to some extent this inconvenient pattern. In the next section, I additionally provide alternative specifications that estimate econometric specifications, excluding the data from October 2019 onwards to avoid likely influential observations from the period of massive social protests in Chile, known as "Social Unrest"[5] as well as the COVID pandemic from March 2020 on. Alternatively, I try to control for these events in different ways. While all transition probabilities show a sharp comeback to pre-pandemic levels, the *oi* probability remained hovering at the slightly lower average level. All informality inflow probabilities, especially the *ui* flow, show a sudden increase in March

[5]Starting on October 18th, 2019, a massive wave of demonstrations, protests, and riots occurred in Chile, disrupting the public transportation system, classes in schools and universities, and many productive activities, especially retail, which suffered looting. The police and the military tried to preserve public order and safety, sometimes exerting excessive repression. The authorities also dictated a night curfew for several days. These events, known in Chile as "Estallido Social", led to substantial effects on economic activity and the functioning of the labor market in Chile during the last quarter of 2019 until the onset of the COVID-19 pandemic.

2020, introducing newcomers to informality and therefore reducing the persistence at this time.

Figure 3 shows the behavior of the series before COVID. The huge changes observed right after March 2020 prevent us from appreciating more subtle trends in the data before COVID. During the period March 2010–February 2020, we can appreciate that all transitions had much more stable behavior. The formal-to-informal weekly transition probability $fi$, shows a mild decline until 2014, while the persistence of informality $ii$, exhibits a mild increase. The unemployment-to-informality transition probability, $ui$, shows a relatively stable evolution since 2010, averaging around 3%. The inflow from inactivity $oi$, exhibits a stable pattern, hovering around 1% per week. DFL weighting barely affects the results, suggesting that compositional issues play a minor role. The increasing persistence until 2019 is associated with a clear but moderate downward trend in informality inflows from formality and inactivity. Newcomers in informality show a slight decrease in their importance, increasing the relative importance of informality stayers.

Figure 4 shows informality outflows at a quarterly frequency, i.e., with no time-aggregation bias correction. Quarterly probability transitions serve as a benchmark to highlight the implications of neglecting the bias. As in the weekly case, DFL adjustments barely affect the series. After a decade of a mild decay of formality-to-informality flow, with an average of around 6.5% per quarter, a sudden spike of nearly 17% occurred in March 2020, followed by a large drop and a subsequent recovery at a new quarterly probability of around 5% on average. The data shows a slight growth in informality-stay probability, $ii$, until the COVID pandemic onset, when this variable plummeted from nearly 55% to almost 10%. During 2020, the $ii$ flow recovered fast to reach pre-pandemic levels around 2022. The informality inflow from unemployment, $ui$, shows a relatively stable trajectory until a few months before the pandemic hit. As mentioned above, in October 2019, there were massive protests and riots in Chile, the "Social Unrest", which may be related to the large reduction preceding the COVID onset from nearly 32% to almost 15%. At the onset of the onset of the COVID pandemic, another large drop took place, hitting almost 5% per quarter. During the rest of 2020, this transition probability increased until it reached levels before the pandemic. Informality inflows from inactivity, $oi$, show a steady decline from approximately 11% in 2010 until 9% by mid 2019. At the end of 2019, there was a huge spike of 13%, probably linked to the "Social Unrest", and then a big drop, almost reaching 5%. During 2020, there was a partial reversal, in which the $oi$ flow hovered around 10%.

To further understand the trajectory of quarterly informality inflows prior to the pandemic, I exhibit Figure 5. The $fi$ transition probability followed a declining trend from 2010 to 2014 and then a minor increasing trend until 2018. From 2010 until the beginning of 2020, the stay informality flow, $ii$, continued its rising trend between 54% and 62%, which imply expected durations of 2.2 to 2.6 quarters, largely greater than those implied by weekly probability transitions. The inflow from unemployment exhibited a decrease until 2014 to remain hovering around 20% until mid-2019. Its volatility seems to be rising over time, with notable spikes exceeding 25% in late 2018 and late 2019. Conversely, for the whole pre-pandemic period, the inflow from inactivity exhibits a relatively stable pattern around 11%.

For the sake of completeness, I display in the appendix contains figures of all transition probabilities. The formality inflows are shown in Figures A1– A4. The inflows into unemployment are depicted in Figures A5–A8. Finally, I plot the inactivity inflows series in Figures A9–A12.

Figure 2: Evolution of weekly informality inflows transition rates

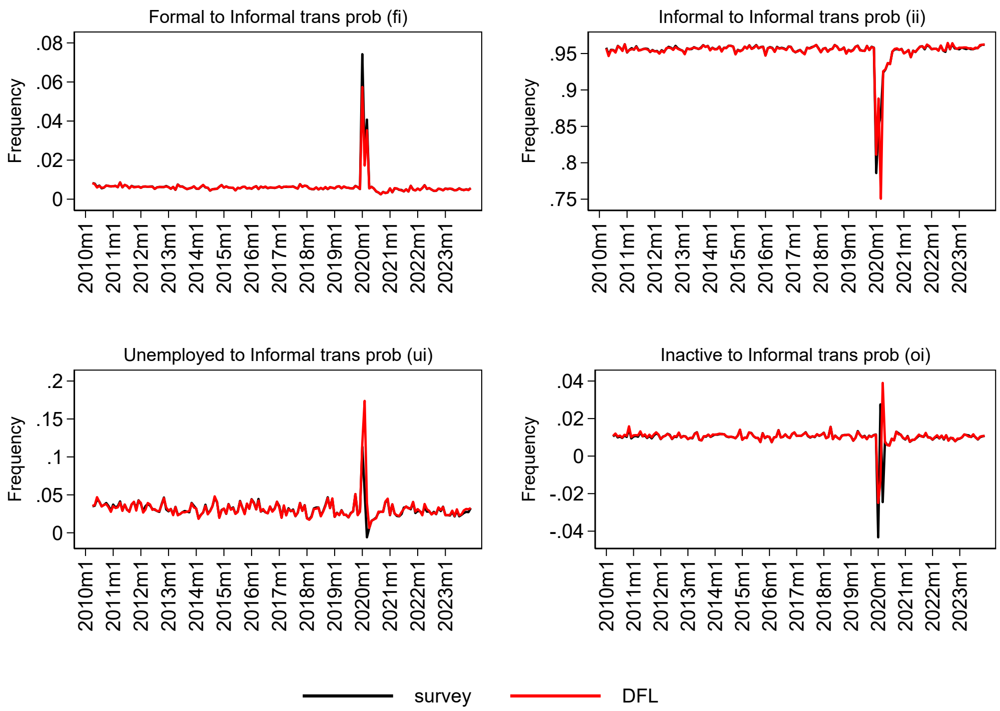



Figure 3: Evolution of weekly informality inflows transition rates pre COVID

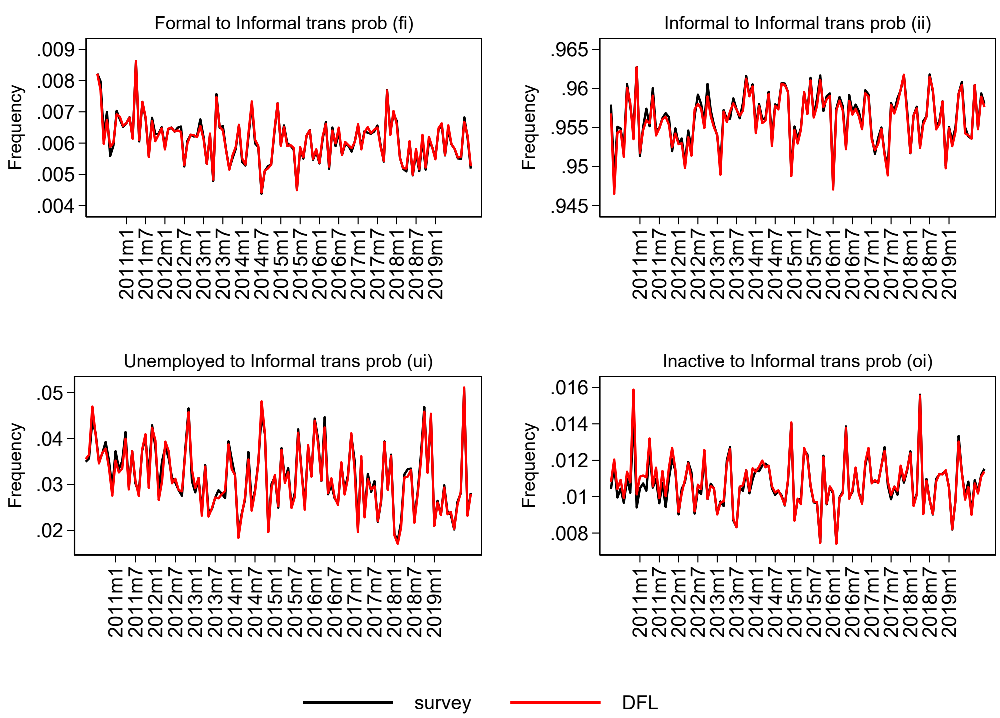


Figure 4: Evolution of quarterly informality inflows transition rates

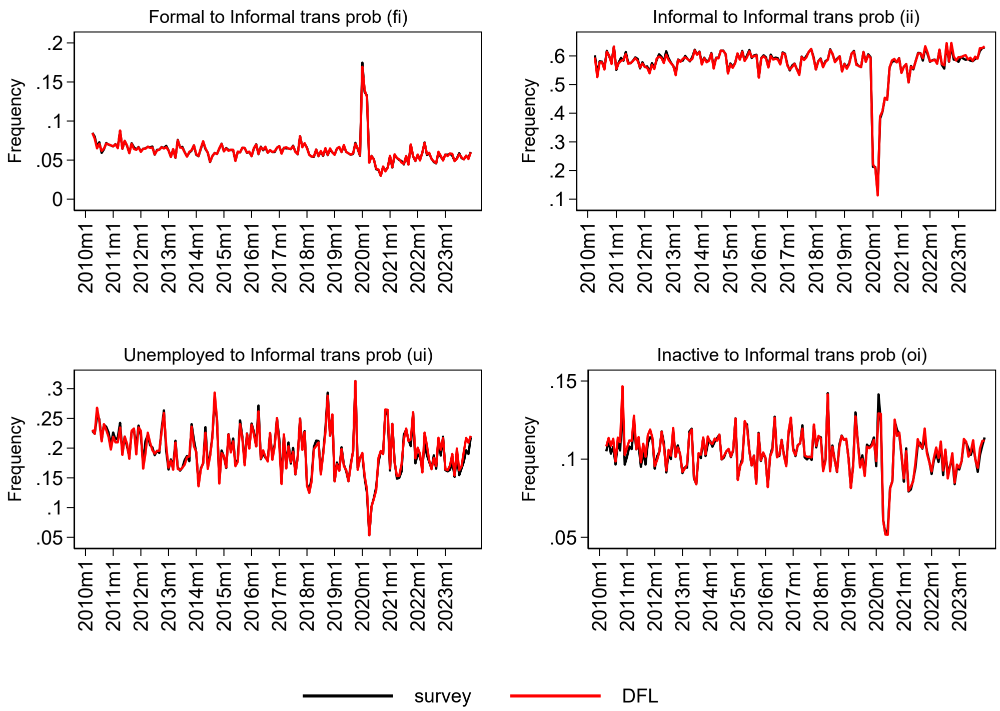
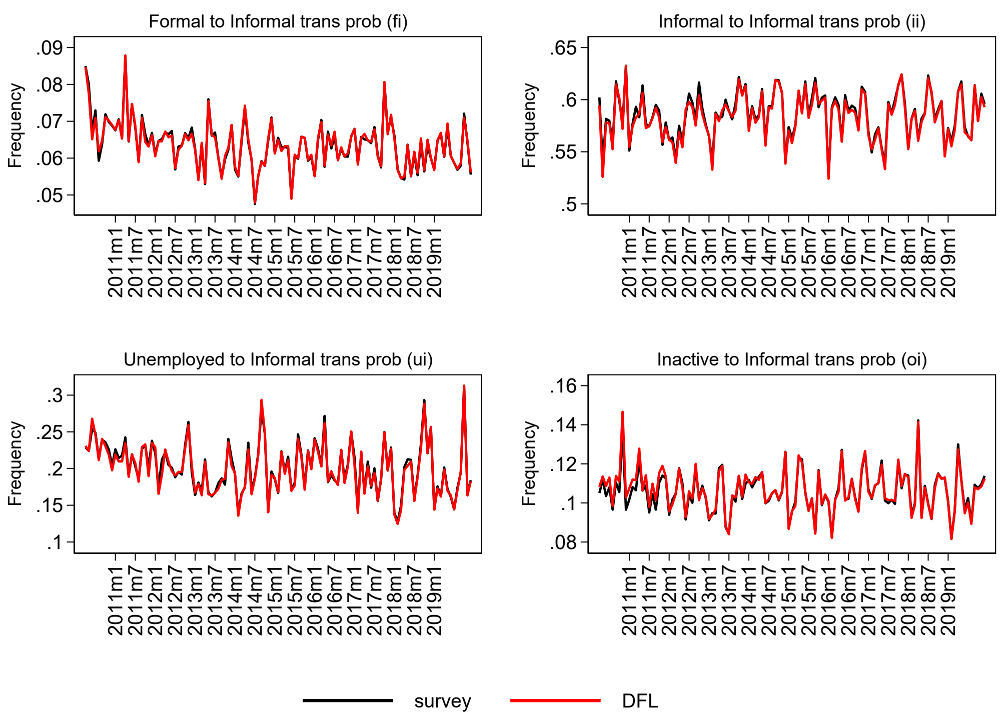



Figure 5: Evolution of quarterly informality inflows transition rates pre COVID



## 4.2 Informality persistence decomposition

I compute the decomposition portrayed in equation (3) using both weekly transition rates, as defined in (4), and quarterly, computed from ENE data using DFL weights, to illustrate the effect of correcting for time-aggregation bias. For weekly and quarterly transitions, the empirical measurement of partial derivatives using local projections in equation (5) requires careful interpretation. For weekly transitions, the equations allow estimating the effect on the "average" weekly informality inflows in the next quarter $t+1$, e.g., $fi$, $ui$, and $oi$ transitions of an increase in today's informality share $I_t$. For quarterly transitions, the estimates measure the effect on the quarterly transition informality inflows in the next quarter $t + 1$ of a raise in the informality share in $t$.

Table 3 shows the main results of the persistence decomposition for DFL-weighted flows and shares. I report three different specifications. In all of them, I control for monthly seasonal binary variables and fixed effects for the categories defining the bins: gender, age group, and education group. In (A) I also control for a COVID-onset binary variable that takes values 1 for March to December 2020 and a "Social Unrest" binary variable from October 2019 onwards. In (B), I only estimated the model using data before October 2019, so Social Unrest or COVID do not affect these results.[6] Besides the controls present in (A), the specification (C) also controls for the shares by gender-age-education bins of immigrant workers, household heads, married- partners, and residence in four geographical zones (Santiago, North, Center excluding Santiago, and South).

The results for weekly frequency in the superior panel are robust across specifications and render an estimate of weekly persistence in the ballpark of 0.935, implying an average duration of $1/(1-0.94) \approx 15.4$ weeks, approximately 3.6 months. The standard deviation is low, so the estimate of the average persistence is quite precise. The decomposition proposed of the persistence quantity $\rho - 1$ in equation (3) shows that the unemployment inflow to informality $U_t \frac{\partial ui}{\partial I_{t-1}} - ui_t$ is the most important reason for workers to get into informality within a week, accounting for 45% of the total for the whole sample and even 57% before October

[6]No anticipation effects are likely important since the COVID pandemic was clearly unexpected. As for the Social Unrest, many opinion leaders and politicians claimed that the inception of massive protests and the simultaneous incendiary attacks on several subway stations in Santiago were unanticipated.

2019, with ample statistical significance. I interpret that informality becomes a way to avoid unemployment, likely a way of self-insurance. The decrease in informality before COVID, even under weak labor market conditions, suggests that the sanitary conditions made informal jobs, usually non-amenable to remote work, less appealing after March 2020, possibly explaining the reduction in importance of the $ui$ flow to shape informality persistence.

The rise of informality-stayers due to an increase in informality share accounts for 25-30% of the whole sample, but only for 9% before October 2019. While the point estimate is large compared to other effects, the standard errors are large, and I could not reject a zero contribution. Transitions to informality from inactivity account for 18% of inflows contribution and up to 22% before October 2019. As many inactive workers are borderline unemployed, the interpretation of this result may go along the same lines as in the case of unemployment. Transitions to informality from formal jobs account for 8% of inflows contribution and nearly 12% before October 2019 and are statistically significant.

The lower panel of Table 3 shows what happens if the time-aggregation bias correction is not applied. I do the same decomposition in (3) using the original quarterly data, weighted by survey and DFL weights. The implied persistence of informality with quarterly data is estimated at around 0.63, which implies an expected duration of $1/(1-0.59) \approx 2.43$ quarters, i.e., 7.3 months, which doubles the duration obtained with weekly transition probabilities. The results vary little across different specifications, including those estimated with data before October 2019, and have a low standard deviation.

Focusing on the decomposition of informality inflows to account for $\rho - 1$, the most important, as in the weekly case, is the $ui$ flow, whose contribution hovers around 43% across specifications but does not change by much when excluding the last years (row B). Transitions from inactivity account for nearly 25%, and the decrease of the informality-stay flow due to the informality increase nearly explains 20% of $\rho - 1$ for the whole sample, but only 9% for the sample until October 2019. Transitions from formality account for 15% of the transitions into informality at a quarterly frequency. Taken together, these pieces of information form two observations. First, the ranking of factors accounting for persistence is roughly the same when using quarterly transitions. Second, the results in both weekly and quarterly transitions are remarkably invariable to the specifications of the models and the sample used.

Table 3: Informality persistence decomposition: DFL weights

| | model | $\rho$ | dur | $I_t \frac{\partial ii}{\partial I_{t-1}}$ | $F_t \frac{\partial fi}{\partial I_{t-1}} - fi$ | $U_t \frac{\partial ui}{\partial I_{t-1}} - ui$ | $O_t \frac{\partial oi}{\partial I_{t-1}} - oi$ |
|---|---|---|---|---|---|---|---|
| weekly frequency | | | | | | | |
| coef | | 0.933 | 14.9 | -0.017 | -0.006 | -0.031 | -0.013 |
| sd | A | (0.009) | | (0.015) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 25.9 | 8.9 | 46.5 | 18.7 |
| coef | | 0.937 | 16.0 | -0.006 | -0.007 | -0.036 | -0.014 |
| sd | B | (0.004) | | (0.006) | (0.001) | (0.002) | (0.002) |
| % $\rho - 1$ | | | | 9.0 | 11.7 | 56.8 | 22.4 |
| coef | | 0.931 | 14.5 | -0.021 | -0.006 | -0.030 | -0.012 |
| sd | C | (0.009) | | (0.016) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 30.8 | 8.1 | 43.4 | 17.7 |
| quarterly frequency | | | | | | | |
| coef | | 0.592 | 2.5 | -0.082 | -0.054 | -0.173 | -0.099 |
| sd | A | (0.009) | | (0.016) | (0.004) | (0.006) | (0.009) |
| % $\rho - 1$ | | | | 20.2 | 13.2 | 42.4 | 24.3 |
| coef | | 0.584 | 2.4 | -0.039 | -0.070 | -0.188 | -0.119 |
| sd | B | (0.013) | | (0.022) | (0.004) | (0.007) | (0.012) |
| % $\rho - 1$ | | | | 9.3 | 16.7 | 45.3 | 28.7 |
| coef | | 0.596 | 2.5 | -0.078 | -0.053 | -0.175 | -0.097 |
| sd | C | (0.009) | | (0.017) | (0.004) | (0.006) | (0.010) |
| % $\rho - 1$ | | | | 19.3 | 13.1 | 43.5 | 24.1 |

Note: Specification A for controls for COVID onset, March to December 2020, and the Social Unrest binary variable from October 2019 onwards. Specification B excludes observations after October 2019. Specification C is like A but includes shares of immigrant workers, household heads, married or partners, and placement in four geographical zones (Santiago, North, Center excluding Santiago, and South). All specifications include fixed effects for bin variables: gender, age group, and education group, as well as monthly seasonal dummies.

What can be learned from contrasting the decomposition of persistence at weekly and quarterly frequencies? Failing to consider time aggregation bias generates estimates of substantially more persistence and longer informality durations on average. Quarterly frequencies of these flows likely mask short informal employment and unemployment spells. For this reason, the result is expected, as most informal jobs require little or no preparation time and workers do not need to go through time-consuming recruiting processes. The contribution of different informality inflows in shaping its persistence is relatively similar in both frequencies. While the persistence does not change significantly focusing solely on data before October 2019, the relative weight for each inflow is somewhat different. The main picture that shows these decomposition exercises is the same: informality duration is short in comparison to estimates available from other Latin American economies (Bosch and Maloney, 2010; Haanwinckel and Soares, 2021). If we consider the sources of persistence, an important "escape route" out of unemployment is through informality, in line with the view of this situation as an adjustment between employment and unemployment (David et al., 2021).

Unfortunately, the rotating panel of the ENE survey is too short to track the medium-term status of individuals who switch from informality to unemployment to observe if they succeed in finding formal jobs on overage. Greater inflows into informality mechanically reduce the importance of stayers in the pool of informal workers. For the persistence to be lower in Chile, it is necessary that inflows into informality, especially $ui$, are relatively large in comparison to other labor markets. These observations suggest the presence of substantial job opportunities in the informal Chilean labor market that constantly attract unemployed workers, although my analysis cannot inform about the long-term state of the worker after a switch.

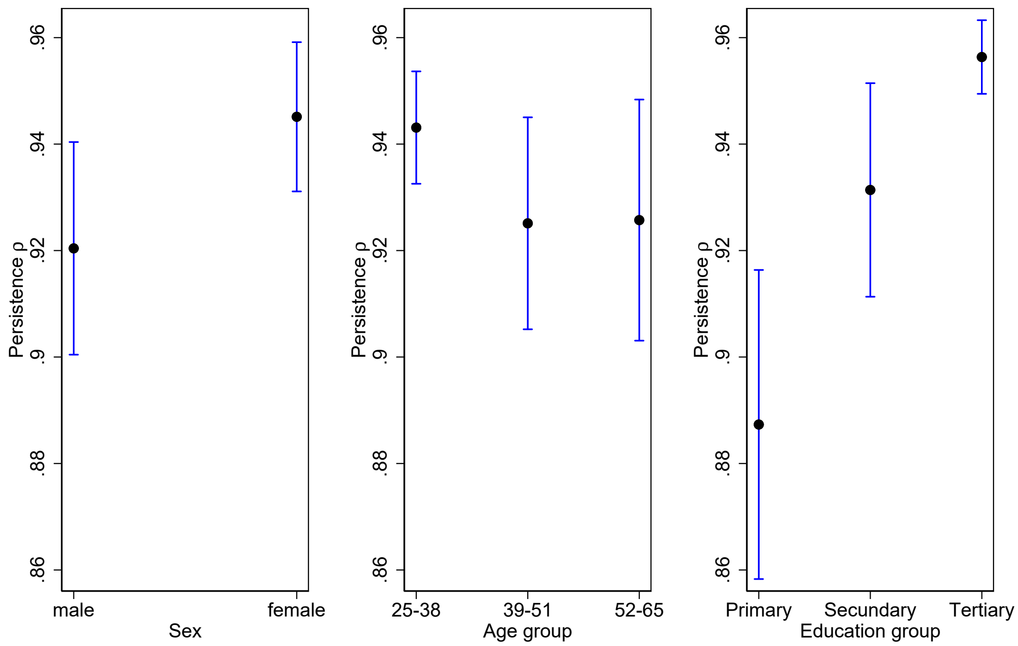
Why is this situation happening in Chile in comparison to other Latin American economies? The most plausible hypothesis is that the informal sector is indeed much smaller in relative terms in Chile, and therefore, substantial inflows affect the informality share more in relative terms. Perry et al. (2007); Maurizio (2012); David et al. (2021) and others point out that the Chilean economy has one of the lowest informality shares in Latin America, which is expected given the development of the country (Donovan et al., 2023). Some jobs for low-skilled workers tend to be covered by formal contracts in Chile, but probably not in other countries. Table 2, in the description of the data, shows that there is a fare share of informal workers even for college workers in occupations with high skill requirements, although there is a clear majority share of formal workers. This suggests that a vast majority of Chilean workers have real job opportunities in both the formal and informal sectors. Informality offers a way to shelter from unemployment for most workers in the economy.

### 4.3 Heterogeneity results

Informality dynamics clearly differ among groups in the population. As stated above, dealing with individual-level transitions impedes correcting for time aggregation bias as we do not observe transition probabilities at that level. Hence, we need to aggregate and rely on the law of large numbers to hopefully correct this issue. The empirical strategy devised in (5) provides a middle ground between aggregation and heterogeneity. To assess the heterogeneous expected persistence between different groups, I plot the conditional estimated probabilities with their corresponding confidence intervals in Figure 6. The left panel plots the confidence intervals and point estimates for persistence conditioning in gender. It is apparent that the persistence of males is lower than that of females, implying that the expected informality duration of the former is shorter. Nevertheless, both confidence intervals overlap, so we cannot reject the fact that the average persistence of both groups is the same at 5% of significance. The middle panel of the figure exhibits the confidence intervals and point estimates by age group. The youngest have higher persistence. Mid-aged individuals (39–51) and older individuals (52–65) have very similar persistence. Nevertheless, the confidence intervals of all age groups overlap, i.e., age persistence gaps are not different in this dimension. Finally, the result of informality persistence differences by education level is perhaps unexpected in that workers with a college degree or more exhibit seem more attached to informality, although Bobba et al. (2022) report a similar finding for Mexico. In contrast, individuals with educational attainment below high school are the ones with lower informality persistence. All these differences suggest that there is important heterogeneity among demographic groups that reflects the quality of job opportunities. More educated workers likely have better informal jobs than the average informal worker. On the other hand, the young and females may face relatively less attractive prospects in the formal labor market, so they remain in informality longer.

I compute the corresponding persistence using the coefficients estimated by local projections in (5) in the synthetic panel across bins and over months. The generated data is weighted by the number of individuals this bin represents to construct some measures of the distribution of persistence over time. Figure 7 exhibits the median of the persistence

Figure 6: Heterogeneous weekly persistence by groups, DFL weights

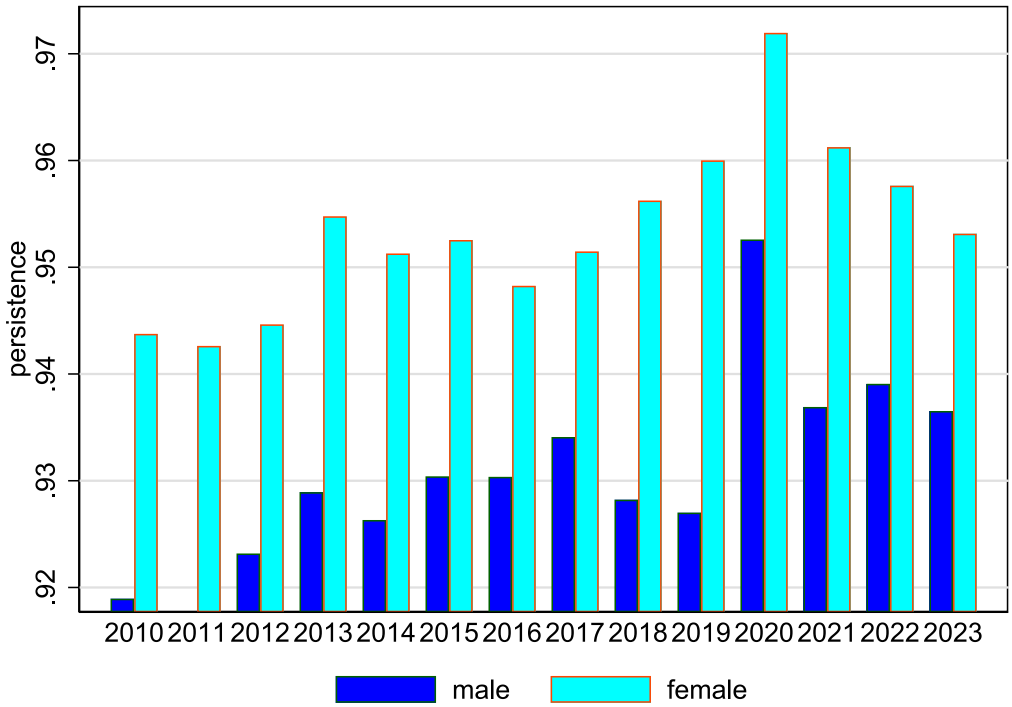



Note: results are based on a specification in which local projections include transition lags and a dummy adjustment for COVID onset, March to December 2020.

distribution conditional on gender and year. Upward trends in the persistence of informality occur until they peak in 2020. In 2021–23, the median persistence decreased but remained somewhat higher than the values observed before 2020. A second salient feature of Figure A20 is a consistent gender gap, with females's informality persistence being higher than that of males.

The informality persistence median trajectories of workers by educational attainment are displayed in Figure 8. In all years, individuals with tertiary educational levels have higher persistence than those reaching secondary, which, in turn, surpasses the persistence of those who completed primary education. For the later group, the persistence is roughly increasing over time, but for other educational groups, the trajectories seem roughly stationary around 0.93 and 0.96 for secondary and tertiary, respectively. In 2020, the persistence of all groups peaked and the gaps among the groups narrowed. Then the series decreased to reach values close to the pre-pandemic levels of persistence. The fact that more educated workers have seemingly more attachment to informal jobs entails the view that work opportunities in the

Figure 7: Median weekly persistence over time by gender, DFL weights




Note: results are based on a specification in which local projections include transition lags and a dummy adjustment for COVID onset, March to December 2020.

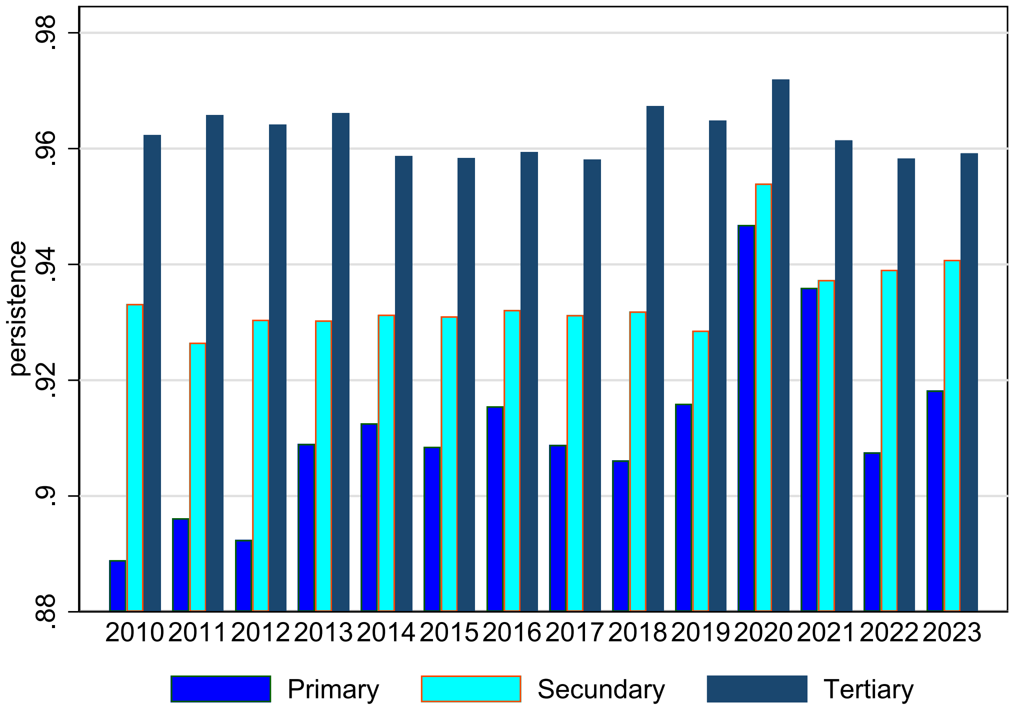
labor market are attractive, especially for individuals with more education and experience (Günther and Launov, 2012).

### 4.4 Robustness checks

I explore different specifications, samples, and weighting schemes to check the robustness of my findings in the preceding section. The detailed results are in the appendix. I am presenting a description of alternative models and discussing the implications.

First, I redo the decomposition using weekly and quarterly transitions with specifications D, E, and F in Table A1. Specification D is similar to A in Table 3, but I control for the COVID pandemic using a dummy variable that takes value 1 from March 2020 on, so it tries to capture COVID as a rather permanent shock while still controlling for Social Unrest, bin variables, fixed effects, and monthly seasonal variables. Specification E tries to control for a secular time trend that eventually changes once COVID hits, i.e., I include a COVID dummy along the lines of specification D, but I also add an interaction term with a linear

Figure 8: Median weekly persistence over time by college status, DFL weights




Note: results are based on a specification in which local projections include transition lags and a dummy adjustment for COVID onset, March to December 2020.

trend and the linear trend itself as controls. Specification F gets rid of all COVID and Social Unrest controls and just keeps the monthly seasonal array of binary variables as controls. The idea here is to compare this result with the restricted-sample specification B to check how relevant the months after October 2019 have been in shaping the measured persistence without taking into account Social Unrest and COVID.

The results in Table A1 show that the results barely change at the weekly frequency, nor does the persistence decomposition. As in the baseline Table 3, the contribution to the persistence quantity $\rho - 1$ of the unemployment inflow into informality, $ui$, is a similar 50%. The persistence in the stay-flow reaches a 25% contribution but is imprecisely estimated. Inactivity inflows account for 18% of persistence. As for quarterly transitions, the estimates are quite similar to one another, reaching 59%. The inflow decomposition is very similar across specifications and close to the values displayed in Table 3.

I allow for a richer dynamic structure in the local projection equation (5) to include lagged values of transition probabilities in quarterly frequency. Report these estimates in

Table A2 for all specifications shown in Tables 3 and A1 (A-F). Overall, the results decrease the estimated persistence in the margin, obtaining point estimates ranging between 0.922 and 0.939. For most specifications, the inclusion of transition probability lags reduced the estimated persistence of informality marginally, rendering expected informality durations between 12.8 and 16.4 weeks. Restricting the sample until September 2019 makes the persistence estimate relatively low, but still in the ballpark of other specifications. The decompositions of persistence among informality inflows exhibit higher variance in this case, but in every case, the most important contribution to persistence is the inflow from unemployment, $ui$. At the quarterly frequency, in the inferior panel of Table A2, the inclusion of lagged transition probabilities caused very little impact in the estimates, rendering an implied average duration of around 2.5 quarters. The informality persistence decomposition looks quite similar across different specifications. As in the weekly frequency case, the $ui$ flow accounts for the largest share in persistence, in the range of 42-45%.

In Table A3, I explore the role of compositional changes in driving results observed by only using the ENE survey weights to construct the flows, which allows for compositional bias. As suggested in the transition series computed with and without the DFL weighting scheme, the results are quite similar, and here happens the same. The persistence estimates roughly remained unaltered, and the same applied to different specifications for weekly and quarterly frequencies. The relative importance of different informality inflows in shaping informality persistence resembles quite closely that estimated in Tables 3 and A1.

Table A4 combines the survey weighting and the introduction of transition probability lags. While the persistence estimates show a marginal reduction in many cases compared to other estimates, the importance of inactivity inflows into informality, i.e., $oi$ flows, slightly increases. All in all, the results show that, regardless of the choices made in modeling the impact of COVID (even getting rid of the data from this period), the weighting scheme, or the dynamic specification in the local projections, the persistence is reasonably robustly estimated, as is the contribution of different inflows into informality.

I also report the estimated evolution of heterogeneous effects by gender, age group, and educational group for weekly estimates in the appendix. I plot additional Figures, including lagged probability transitions at the local projections, in Figures A17, A20, and A23. I also considered the effect of not accounting for composition by using ENE survey weights in Figures A18, A24, and A21. I additionally included the results using both lagged probability transitions and ENE survey weights in Figures A19, A22, and A25. While there are some differences in estimates generated by different assumptions, the main features are those described in Section 4.3.

## 5 Conclusions

Researchers have mainly focused on understanding the empirical determinants of informality, adopting a static perspective. Earlier literature often adopted the view that the formal and informal sectors are separated. A modern view, empirically justified by the study of worker flows and theoretically amenable to search and matching models, allows us to understand a dynamic labor market in which individuals transition to and from informality. The causes of being informal are undoubtedly important, but the evidence shows that informality is often temporal, especially in Chile. This is not to say that society and policymakers should not care about this: informality matters because it is often linked to low wages and risky working conditions and has strong implications for welfare and inequality. Moreover, informal firms can often survive, despite their low productivity, by skipping regulations and avoiding taxes that are costly at the micro level, perpetuating poor managerial practices, scant technological adoption, and innovation (La Porta and Shleifer, 2014).

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 chain approximation and use a time-aggregation bias correction, as in many other papers dealing with transition probabilities in the labor market (Shimer, 2012; Elsby et al., 2015; Choi et al., 2015; Gomes, 2015; Bosch and Maloney, 2010). I also use the DiNardo et al. (1996) technique to reweight and correct for compositional changes in the rotation panel of the employment survey. To account for the dynamic impact of stocks on transition probabilities, I use local projections (Jordà, 2005).

I find that the measured average informality persistence in Chile seems substantially lower compared to the available evidence for big Latin American economies: Argentina, Brazil, and Mexico (Bosch and Maloney, 2010) and Brazil Haanwinckel and Soares (2021).

The quantitative persistence I estimate importantly depends on the time-aggregation: 3.6 months of expected informality spell duration when correction is applied versus 7.3 months when using quarterly raw uncorrected transition probabilities. While the correction is not without caveats, the results are robust to different ways of controlling for the Social Unrest (from October 2019 onwards) and the COVID pandemic (from March 2020 onwards). If we restrict the sample to data before October 2019, the persistence is very similar, too.

Using the decomposition, I show the importance of different inflows to informality to account for their persistence. At both weekly and quarterly frequency, the specifications show that around 45% of the persistence $\rho - 1$ is accounted for by unemployment-to-informality transitions, $ui$, although the contribution reaches 57% for the sample before Social Unrest and COVID. The effect of increasing informality stock and transitions to informality from inactivity accounts for approximately 20% on average across specifications, but it is only significant at quarterly frequency and is substantially lower before October 2019. The contribution of inactivity inflows is significant and remains in the ballpark of 20%, being slightly higher at quarterly frequency. Formality-to-informality transitions, $fi$ flows, contribute around 10% at weekly frequency, and somewhat more when I use quarterly transitions. The analysis also shows some interesting patterns of informality persistence: it has been increasing over time, with a large spike in 2020 and then a reversal afterwards. Tertiary-educated, female, and young individuals tend to have more persistent informality. The finding in relation to education suggests that skilled workers in the informal sector can still have good jobs, so they tend to stay informal more permanently.

While this analysis cannot answer the medium- or long-run labor status of workers due to the limited time dimension of the ENE rotating panel, these results suggest that the unemployed in Chile actively find informal opportunities. The probability flows suggest that there is a vivid transition between informal and formal jobs. Even if the time-aggregation adjustment is disputed due to their underlying assumptions and its underperformance to account for the Social Unrest and COVID turmoils in the labor market, the estimation using unadjusted quarterly frequency transition probabilities still shows substantially lower persistence compared to those reported in other Latin American economies (Bosch and Maloney, 2010). Given that the persistence of informality is relatively low and tends to be higher for skilled workers, one should be cautious about setting the reduction of informality as a goal of public policies. Since informality receives substantial inflows from unemployment, it can be considered a temporary lifesaver of income while workers search for better opportunities in the market.

## 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.

Bobba, M., L. Flabbi, and S. Levy (2022). Labor Market Search, Informality, and Schooling Investments. *International Economic Review 63*(1), 211–259.

Bosch, M. and J. Esteban-Pretel (2012). Job creation and job destruction in the presence of informal markets. *Journal of Development Economics 98*(2), 270–286.

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.

Danquah, M., S. Schotte, and K. Sen (2021). Informal work in sub-Saharan Africa: Dead end or stepping-stone? *IZA Journal of Development and Migration 12*(1), –.

David, A. C., F. Lambert, and F. Toscani (2021). *Informality and labor market dynamics in Latin America*, Chapter 5, pp. 145–166. The Global Informal Workforce: Priorities for Inclusive Growth. International Monetary Fund.

Dell'Anno, R. (2022). Theories and definitions of the informal economy: A survey. *Journal of Economic Surveys 36*(5), 1610–1643.

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.

Donovan, K., W. J. Lu, and T. Schoellman (2023). Labor Market Dynamics and Development*. *The Quarterly Journal of Economics 138*(4), 2287–2325.

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.

Feng, S. and Y. Hu (2013, September). Misclassification Errors and the Underestimation of the US Unemployment Rate. *American Economic Review 103*(2), 1054–70.

Fernández, A. and F. Meza (2015). Informal employment and business cycles in emerging economies: The case of mexico. *Review of Economic Dynamics 18*(2), 381–405.

Fortin, N., T. Lemieux, and S. Firpo (2011). Decomposition Methods in Economics. *Handbook of labor economics 4*, 1–102.

Gomes, P. (2015). The importance of frequency in estimating labour market transition rates. *IZA Journal of Labor Economics 4*, 1–10.

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.

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.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American economic review 95*(1), 161–182.

Kaitz, H. B. (1970). Analyzing the length of spells of unemployment. *Monthly Labor Review 93*, 11.

La Porta, R. and A. Shleifer (2014). Informality and development. *Journal of economic perspectives 28*(3), 109–126.

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.

Maurizio, R. (2012). Labour informality in Latin America: the case of Argentina, Chile, Brazil and Peru. Working Paper 165, Brooks World Poverty Institute Working Paper.

Mondragón-Vélez, C., X. Peña, D. Wills, and A. Kugler (2010). Labor market rigidities and informality in Colombia. *Economía 11*(1), 65–101.

Perry, G. (1972). Unemployment flows in the us labor market. *Brookings Papers on Economic Activity 1972*(2), 245–292.

Perry, G., W. F. Maloney, O. Arias, P. Fajnzylber, A. Mason, and J. Saavedra-Chanduvi (2007). *Informality: Exit and exclusion.* World Bank Publications.

Pissarides, C. (2000). *Equilibrium Unemployment Theory* (Second ed.). The MIT Press.

Poterba, J. M. and L. H. Summers (1986). Reporting errors and labor market dynamics. *Econometrica 54*(6), pp. 1319–1338.

Shimer, R. (2005). The Cyclical Behavior of Equilibrium Unemployment and Vacancies. *American Economic Review 95*(1), 25–49.

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

Tansel, A. and Z. A. Ozdemir (2019). Transitions across labor market states including formal/informal division in egypt. *Review of Development Economics 23*(4), 1674–1695.

Ulyssea, G. (2020). Informality: Causes and consequences for development. *Annual Review of Economics 12*, 525–546.

Verbeek, M. (2008). Pseudo-Panels and Repeated Cross-Sections. In *The Econometrics of Panel Data*, pp. 369–383. Springer.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association 57*(298), 348–368.

# A Online Appendix – not intended for publication

## 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` (gender) 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.

### A.2 Evolution of stocks and transition rates

Figure A1: Evolution of weekly formality inflows transition rates

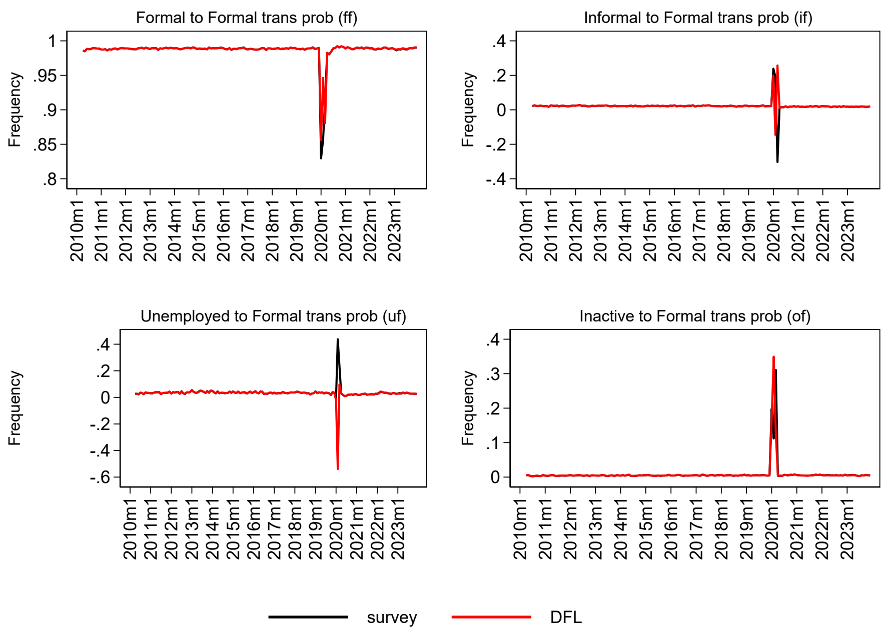



Figure A2: Evolution of weekly formality inflows transition rates pre COVID

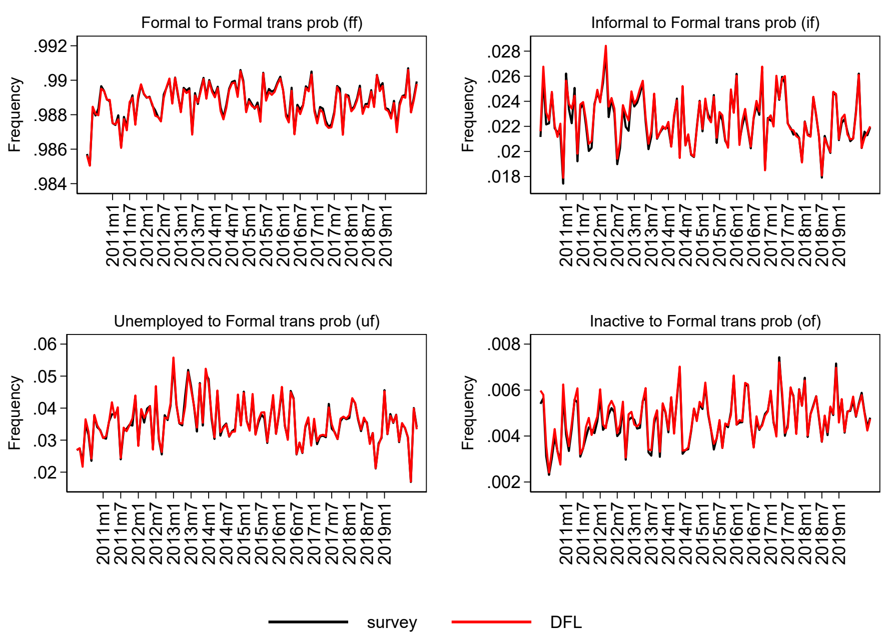


Figure A3: Evolution of quarterly formality inflows transition rates

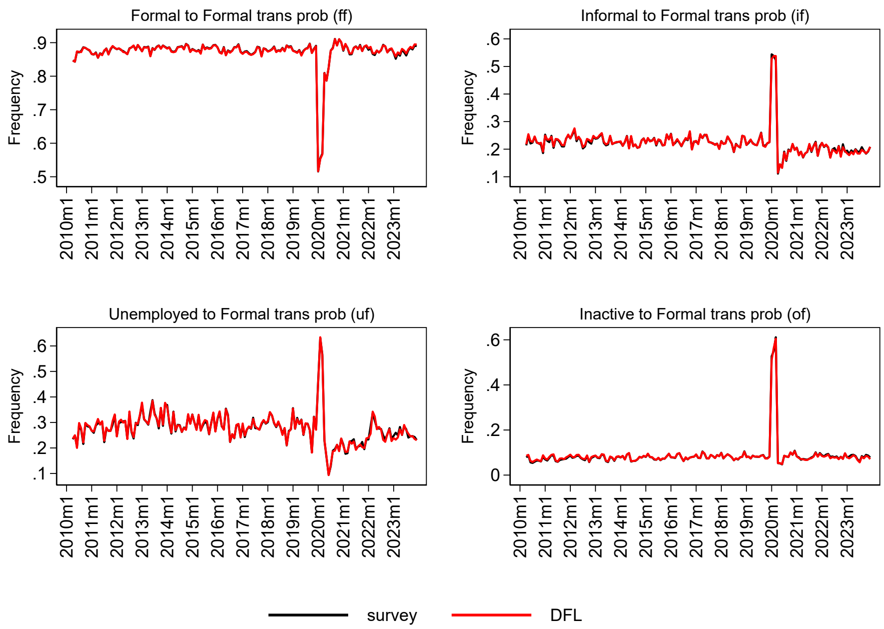



Figure A4: Evolution of quarterly formality inflows transition rates pre COVID

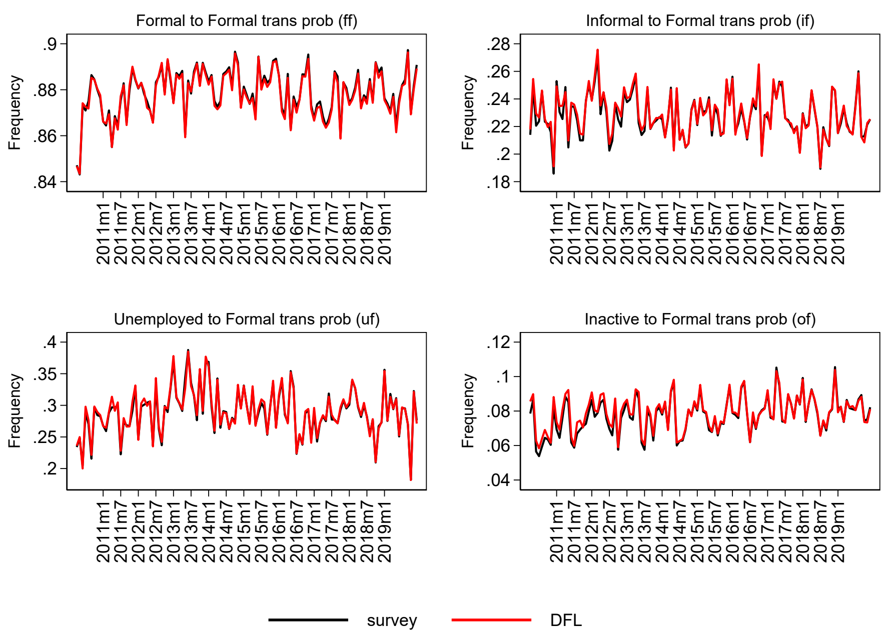


Figure A5: Evolution of weekly unemployment inflows transition rates


Figure A6: Evolution of weekly unemployment inflows transition rates pre COVID

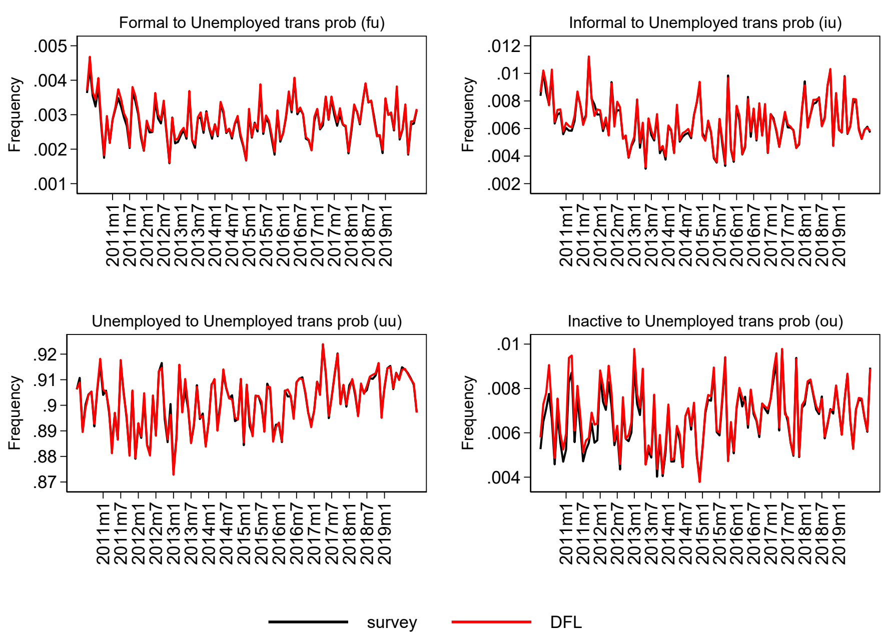


Figure A7: Evolution of quarterly unemployment inflows transition rates

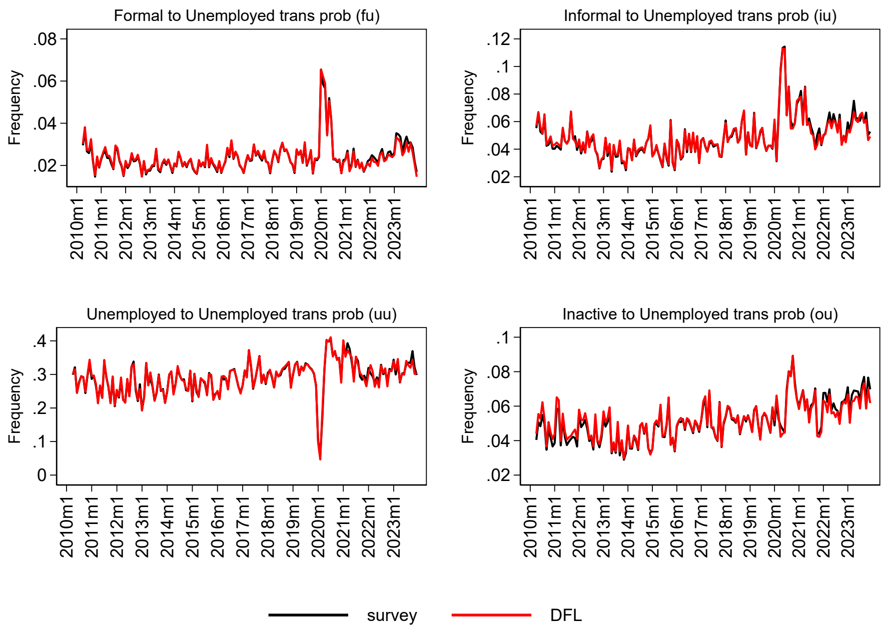



Figure A8: Evolution of quarterly unemployment inflows transition rates pre COVID

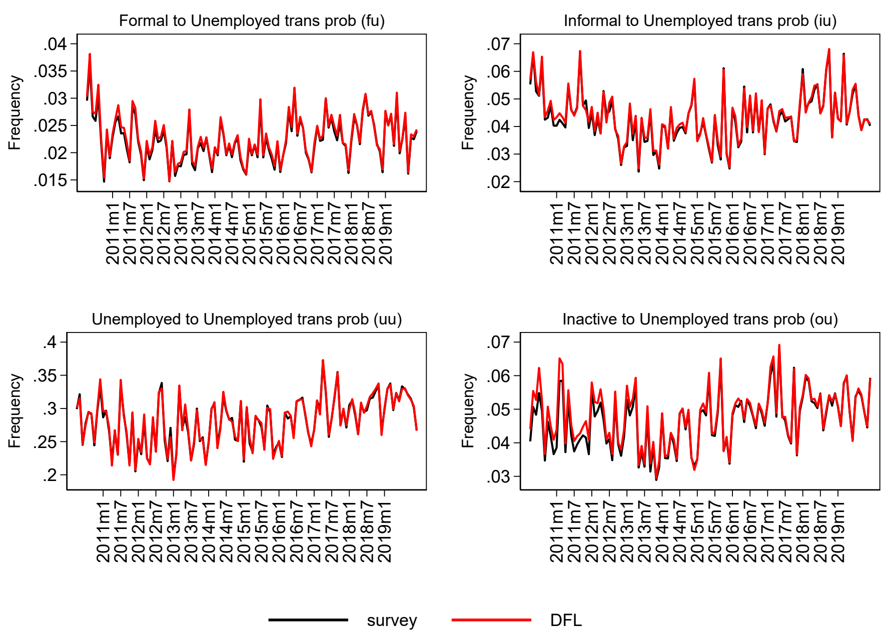


Figure A9: Evolution of weekly inactivity inflows transition rates

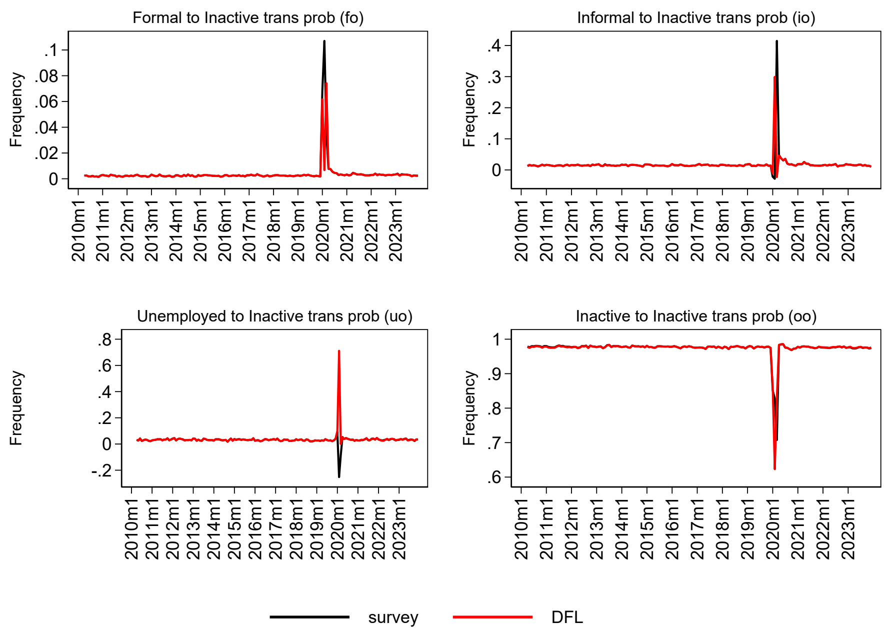



Figure A10: Evolution of weekly inactivity inflows transition rates pre COVID

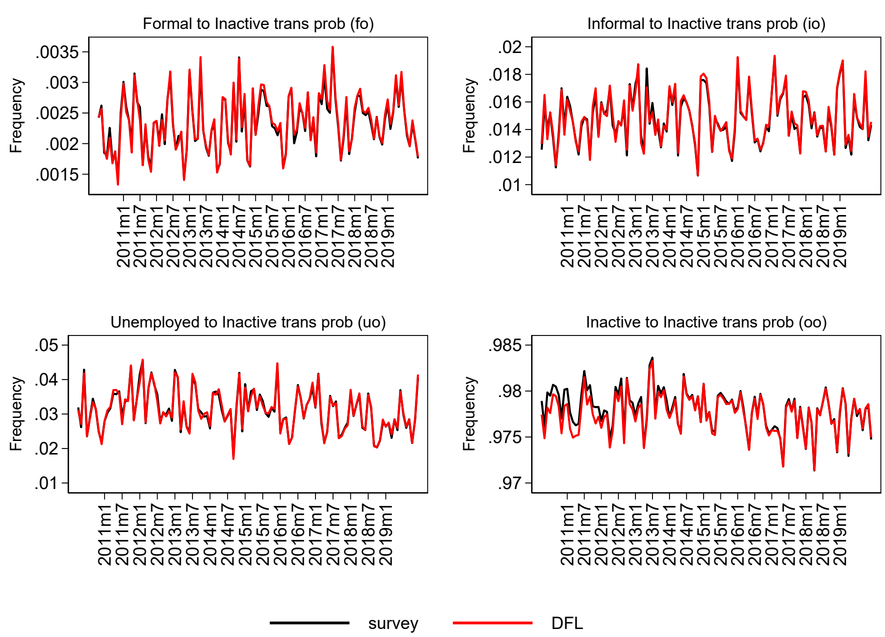


Figure A11: Evolution of quarterly inactivity inflows transition rates

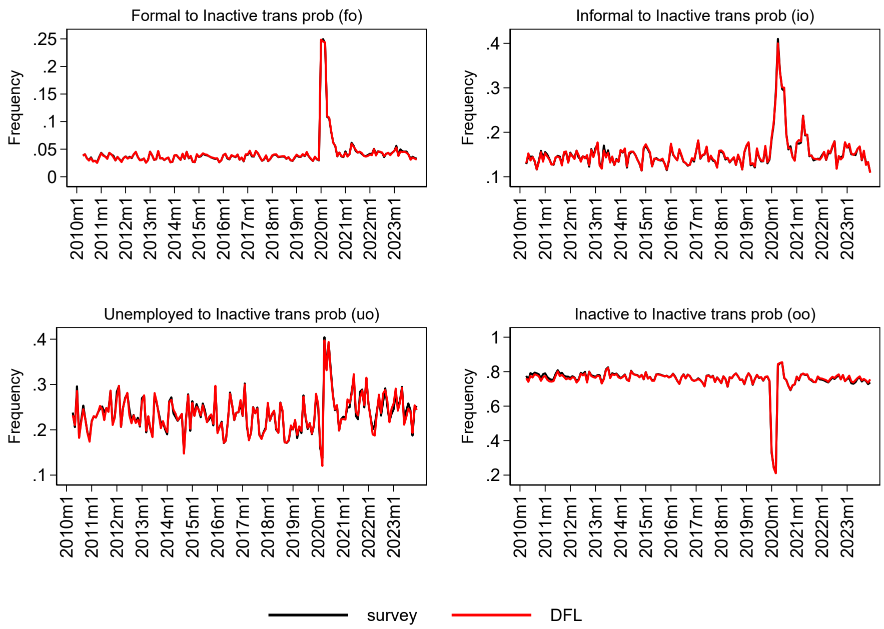



Figure A12: Evolution of quarterly inactivity inflows transition rates pre COVID

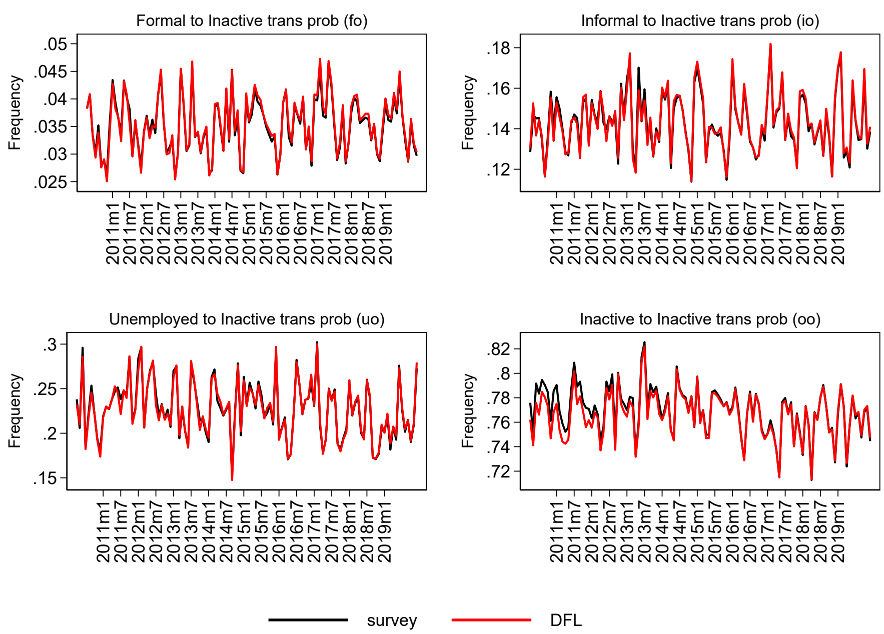


Figure A13: Weekly formality inflows transition probabilities, original and smoothed series

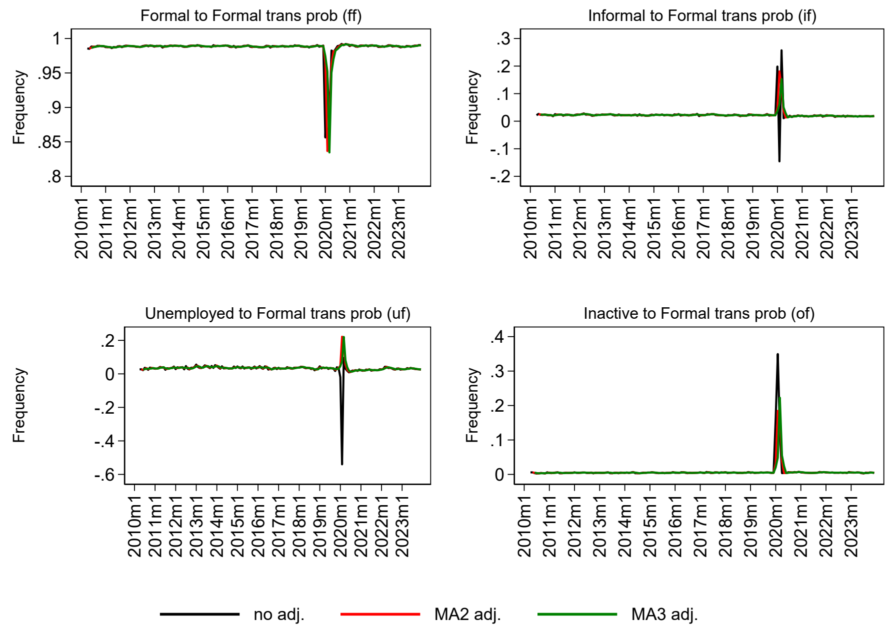



Figure A14: Weekly informality inflows transition probabilities, original and smoothed series

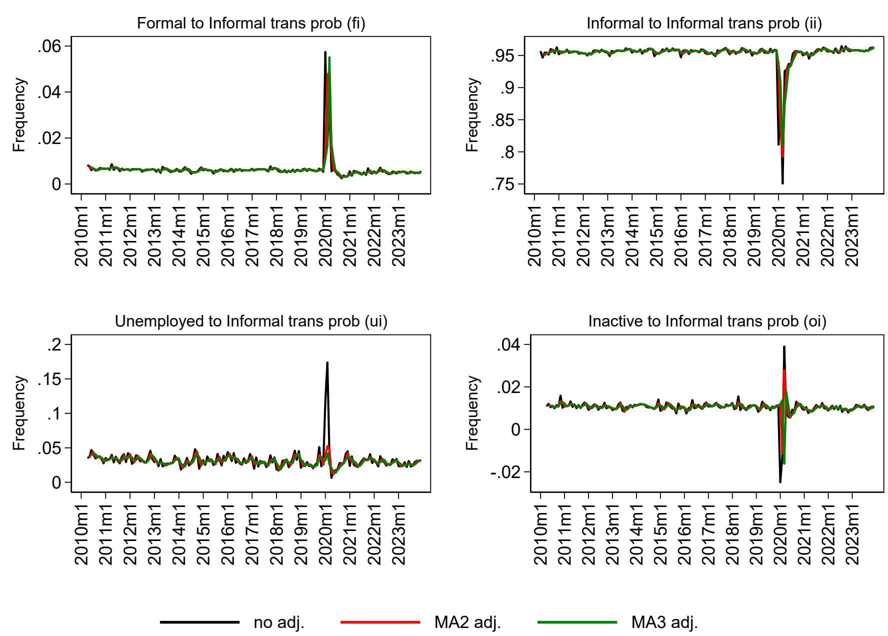


Figure A15: Weekly unemployment inflows transition probabilities, original and smoothed series

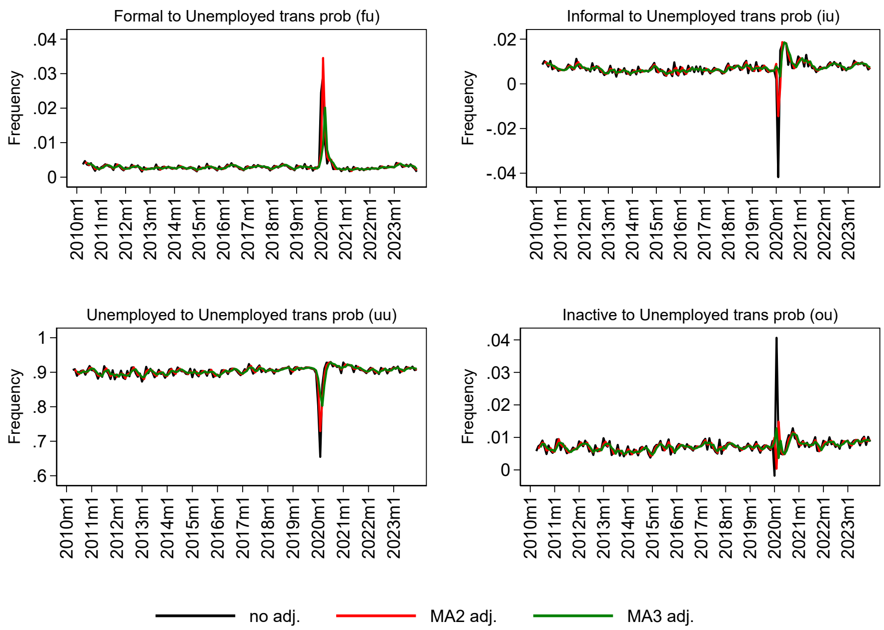
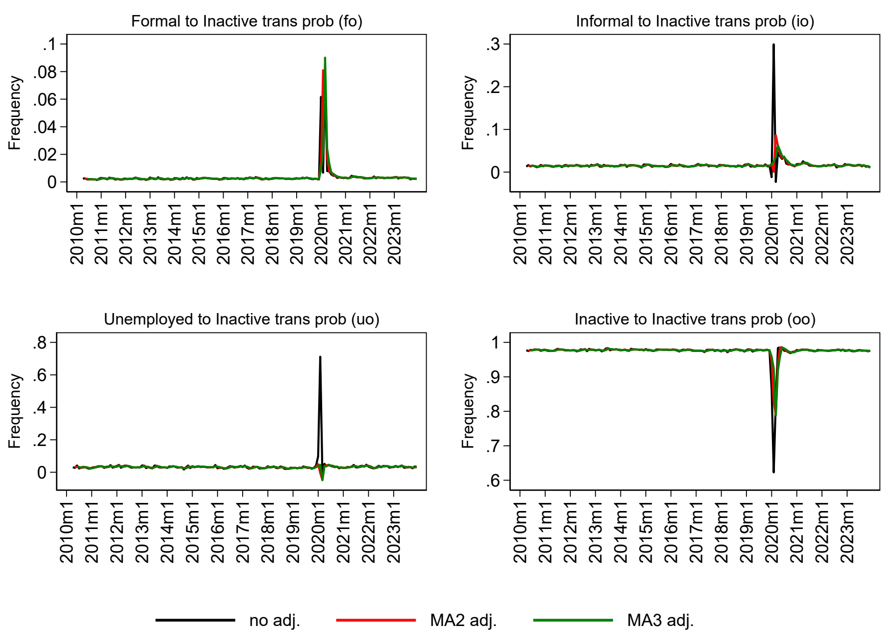



Figure A16: Weekly inactivity inflows transition probabilities, original and smoothed series



## A.3 Additional persistence decomposition results

Table A1: Informality persistence decomposition: DFL weights, other specifications

| | model | $\rho$ | dur | $I_t \frac{\partial ii}{\partial I_{t-1}}$ | $F_t \frac{\partial fi}{\partial I_{t-1}} - fi$ | $U_t \frac{\partial ui}{\partial I_{t-1}} - ui$ | $O_t \frac{\partial oi}{\partial I_{t-1}} - oi$ |
|---|---|---|---|---|---|---|---|
| weekly frequency | | | | | | | |
| coef | | 0.935 | 15.3 | -0.014 | -0.006 | -0.033 | -0.012 |
| sd | D | (0.009) | | (0.015) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 22.0 | 9.6 | 50.0 | 18.4 |
| coef | | 0.935 | 15.3 | -0.016 | -0.005 | -0.032 | -0.011 |
| sd | E | (0.009) | | (0.015) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 24.8 | 8.4 | 49.5 | 17.3 |
| coef | | 0.933 | 15.0 | -0.017 | -0.006 | -0.032 | -0.012 |
| sd | F | (0.009) | | (0.015) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 25.4 | 8.8 | 47.2 | 18.7 |
| quarterly frequency | | | | | | | |
| coef | | 0.592 | 2.5 | -0.082 | -0.054 | -0.173 | -0.099 |
| sd | D | (0.009) | | (0.016) | (0.004) | (0.006) | (0.009) |
| % $\rho - 1$ | | | | 20.2 | 13.2 | 42.4 | 24.3 |
| coef | | 0.584 | 2.4 | -0.039 | -0.070 | -0.188 | -0.119 |
| sd | E | (0.013) | | (0.022) | (0.004) | (0.007) | (0.012) |
| % $\rho - 1$ | | | | 9.3 | 16.7 | 45.3 | 28.7 |
| coef | | 0.592 | 2.5 | -0.085 | -0.049 | -0.173 | -0.100 |
| sd | F | (0.009) | | (0.016) | (0.004) | (0.006) | (0.010) |
| % $\rho - 1$ | | | | 20.9 | 12.1 | 42.4 | 24.6 |

Note: Specification D for controls for COVID onset, March 2020 to December 2023, and Social Unrest binary variable from October 2019 onwards. Specification E includes a binary control taking value 1 for March 2020-December 2020, a linear trend, and their interaction. Specification F gets rid of all COVID variables and linear trends for the whole sample. All specifications include fixed effects for bin variables: gender, age group, and education group, as well as monthly seasonal dummies.

Table A2: Informality persistence decomposition: DFL weights, flow lags

| | model | $\rho$ | dur | $I_t \frac{\partial ii}{\partial I_{t-1}}$ | $F_t \frac{\partial fi}{\partial I_{t-1}} - fi$ | $U_t \frac{\partial ui}{\partial I_{t-1}} - ui$ | $O_t \frac{\partial oi}{\partial I_{t-1}} - oi$ |
|---|---|---|---|---|---|---|---|
| weekly frequency | | | | | | | |
| coef | | 0.929 | 14.0 | -0.018 | -0.006 | -0.029 | -0.017 |
| sd | A | (0.009) | | (0.016) | (0.001) | (0.006) | (0.002) |
| % $\rho-1$ | | | | 25.3 | 9.1 | 41.2 | 24.4 |
| coef | | 0.925 | 13.4 | -0.023 | -0.007 | -0.027 | -0.019 |
| sd | B | (0.009) | | (0.016) | (0.001) | (0.006) | (0.002) |
| % $\rho-1$ | | | | 30.2 | 8.7 | 36.3 | 24.8 |
| coef | | 0.929 | 14.0 | -0.019 | -0.006 | -0.029 | -0.017 |
| sd | C | (0.009) | | (0.016) | (0.001) | (0.006) | (0.002) |
| % $\rho-1$ | | | | 26.0 | 9.0 | 41.1 | 23.9 |
| coef | | 0.939 | 16.4 | -0.003 | -0.007 | -0.037 | -0.014 |
| sd | D | (0.005) | | (0.007) | (0.001) | (0.002) | (0.003) |
| % $\rho-1$ | | | | 5.5 | 11.9 | 60.1 | 22.5 |
| coef | | 0.927 | 13.6 | -0.021 | -0.006 | -0.028 | -0.018 |
| sd | E | (0.009) | | (0.016) | (0.001) | (0.006) | (0.002) |
| % $\rho-1$ | | | | 29.2 | 8.4 | 38.3 | 24.1 |
| coef | | 0.922 | 12.8 | -0.028 | -0.006 | -0.025 | -0.018 |
| sd | F | (0.009) | | (0.016) | (0.001) | (0.007) | (0.002) |
| % $\rho-1$ | | | | 36.4 | 7.9 | 32.2 | 23.5 |
| quarterly frequency | | | | | | | |
| coef | | 0.592 | 2.5 | -0.082 | -0.054 | -0.173 | -0.099 |
| sd | A | (0.009) | | (0.016) | (0.004) | (0.006) | (0.009) |
| % $\rho-1$ | | | | 20.2 | 13.2 | 42.4 | 24.3 |
| coef | | 0.584 | 2.4 | -0.039 | -0.070 | -0.188 | -0.119 |
| sd | B | (0.013) | | (0.022) | (0.004) | (0.007) | (0.012) |
| % $\rho-1$ | | | | 9.3 | 16.7 | 45.3 | 28.7 |
| coef | | 0.592 | 2.5 | -0.085 | -0.049 | -0.173 | -0.100 |
| sd | C | (0.009) | | (0.016) | (0.004) | (0.006) | (0.010) |
| % $\rho-1$ | | | | 20.9 | 12.1 | 42.4 | 24.6 |
| coef | | 0.592 | 2.5 | -0.082 | -0.054 | -0.173 | -0.099 |
| sd | D | (0.009) | | (0.016) | (0.004) | (0.006) | (0.009) |
| % $\rho-1$ | | | | 20.2 | 13.2 | 42.4 | 24.3 |
| coef | | 0.584 | 2.4 | -0.039 | -0.070 | -0.188 | -0.119 |
| sd | E | (0.013) | | (0.022) | (0.004) | (0.007) | (0.012) |
| % $\rho-1$ | | | | 9.3 | 16.7 | 45.3 | 28.7 |
| coef | | 0.592 | 2.5 | -0.085 | -0.049 | -0.173 | -0.100 |
| sd | F | (0.009) | | (0.016) | (0.004) | (0.006) | (0.010) |
| % $\rho-1$ | | | | 20.9 | 12.1 | 42.4 | 24.6 |

Note: Specification A for controls for COVID onset, March to December 2020, and Social Unrest binary variable from October 2019 onwards. Specification B excludes observations after October 2019. Specification C is like A but includes shares of immigrant workers, household heads, married- partners, and placement in four geographical zones (Santiago, North, Center excluding Santiago, and South). Specification D controls for COVID, March 2020 to December 2023, and Social Unrest binary variable from October 2019 onwards. Specification E includes a binary control taking value 1 for March 2020-December 2020, a linear trend, and their interaction. Specification F gets rid of all COVID variables and linear trends for the whole sample. All specifications include fixed effects for bin variables: gender, age group, and education group, as well as monthly seasonal dummies.

Table A3: Informality persistence decomposition: survey weights, no flow lags

| | model | $\rho$ | dur | $I_t \frac{\partial ii}{\partial I_{t-1}}$ | $F_t \frac{\partial fi}{\partial I_{t-1}} - fi$ | $U_t \frac{\partial ui}{\partial I_{t-1}} - ui$ | $O_t \frac{\partial oi}{\partial I_{t-1}} - oi$ |
|---|---|---|---|---|---|---|---|
| weekly frequency | | | | | | | |
| coef | | 0.935 | 15.3 | -0.016 | -0.005 | -0.032 | -0.013 |
| sd | A | (0.008) | | (0.014) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 24.0 | 8.3 | 48.6 | 19.2 |
| coef | | 0.939 | 16.5 | -0.001 | -0.008 | -0.037 | -0.015 |
| sd | B | (0.004) | | (0.006) | (0.001) | (0.002) | (0.002) |
| % $\rho - 1$ | | | | 1.9 | 12.6 | 60.8 | 24.7 |
| coef | | 0.931 | 14.6 | -0.022 | -0.005 | -0.029 | -0.012 |
| sd | C | (0.009) | | (0.015) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 31.9 | 7.1 | 42.9 | 18.1 |
| coef | | 0.937 | 15.8 | -0.012 | -0.006 | -0.033 | -0.012 |
| sd | D | (0.008) | | (0.014) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 19.5 | 9.4 | 52.1 | 19.0 |
| coef | | 0.936 | 15.6 | -0.015 | -0.005 | -0.032 | -0.011 |
| sd | E | (0.008) | | (0.015) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 23.8 | 8.4 | 50.5 | 17.4 |
| coef | | 0.935 | 15.4 | -0.015 | -0.005 | -0.032 | -0.012 |
| sd | F | (0.008) | | (0.014) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 23.5 | 8.3 | 49.1 | 19.1 |
| quarterly frequency | | | | | | | |
| coef | | 0.595 | 2.5 | -0.076 | -0.053 | -0.174 | -0.102 |
| sd | A | (0.009) | | (0.016) | (0.004) | (0.006) | (0.009) |
| % $\rho - 1$ | | | | 18.8 | 13.0 | 42.9 | 25.3 |
| coef | | 0.584 | 2.4 | -0.027 | -0.073 | -0.190 | -0.126 |
| sd | B | (0.013) | | (0.022) | (0.004) | (0.007) | (0.012) |
| % $\rho - 1$ | | | | 6.4 | 17.5 | 45.7 | 30.4 |
| coef | | 0.595 | 2.5 | -0.080 | -0.048 | -0.174 | -0.102 |
| sd | C | (0.009) | | (0.017) | (0.004) | (0.006) | (0.010) |
| % $\rho - 1$ | | | | 19.9 | 11.9 | 43.1 | 25.1 |
| coef | | 0.595 | 2.5 | -0.074 | -0.054 | -0.174 | -0.104 |
| sd | D | (0.009) | | (0.016) | (0.003) | (0.006) | (0.009) |
| % $\rho - 1$ | | | | 18.2 | 13.3 | 43.0 | 25.6 |
| coef | | 0.594 | 2.5 | -0.071 | -0.056 | -0.174 | -0.106 |
| sd | E | (0.009) | | (0.017) | (0.004) | (0.006) | (0.010) |
| % $\rho - 1$ | | | | 17.4 | 13.7 | 42.9 | 26.1 |
| coef | | 0.595 | 2.5 | -0.075 | -0.053 | -0.174 | -0.103 |
| sd | F | (0.009) | | (0.016) | (0.004) | (0.006) | (0.009) |
| % $\rho - 1$ | | | | 18.5 | 13.2 | 42.9 | 25.4 |

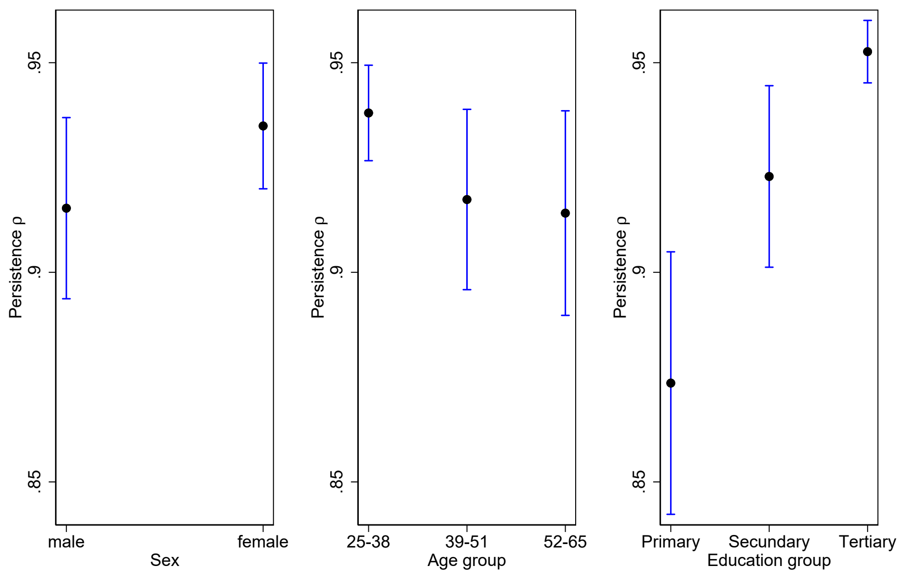
Note: Note: Specification A for controls for COVID onset, March to December 2020, and Social Unrest binary variable from October 2019 onwards. Specification B excludes observations after October 2019. Specification C is like A but includes shares of immigrant workers, household heads, married- partners, and placement in four geographical zones (Santiago, North, Center excluding Santiago, and South). Specification D controls for COVID, March 2020 to December 2023, and Social Unrest binary variable from October 2019 onwards. Specification E includes a binary control taking value 1 for March 2020-December 2020, a linear trend, and their interaction. Specification F gets rid of all COVID variables and linear trends for the whole sample. All specifications include fixed effects for bin variables: gender, age group, and education group, as well as monthly seasonal dummies.

Table A4: Informality persistence decomposition: survey weights, flow lags

| | model | $\rho$ | dur | $I_t \frac{\partial ii}{\partial I_{t-1}}$ | $F_t \frac{\partial fi}{\partial I_{t-1}} - fi$ | $U_t \frac{\partial ui}{\partial I_{t-1}} - ui$ | $O_t \frac{\partial oi}{\partial I_{t-1}} - oi$ |
|---|---|---|---|---|---|---|---|
| weekly frequency | | | | | | | |
| coef | | 0.931 | 14.6 | -0.014 | -0.006 | -0.031 | -0.018 |
| sd | A | (0.009) | | (0.015) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 20.7 | 8.3 | 44.7 | 26.3 |
| coef | | 0.928 | 13.9 | -0.019 | -0.005 | -0.029 | -0.019 |
| sd | B | (0.009) | | (0.015) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 26.9 | 7.3 | 39.8 | 26.1 |
| coef | | 0.931 | 14.5 | -0.015 | -0.006 | -0.030 | -0.018 |
| sd | C | (0.009) | | (0.015) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 22.0 | 8.3 | 44.1 | 25.6 |
| coef | | 0.940 | 16.8 | 0.001 | -0.008 | -0.038 | -0.015 |
| sd | D | (0.005) | | (0.007) | (0.001) | (0.002) | (0.003) |
| % $\rho - 1$ | | | | -1.2 | 12.9 | 63.2 | 25.1 |
| coef | | 0.929 | 14.2 | -0.018 | -0.005 | -0.029 | -0.018 |
| sd | E | (0.009) | | (0.015) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 25.4 | 7.3 | 41.5 | 25.7 |
| coef | | 0.925 | 13.4 | -0.025 | -0.005 | -0.027 | -0.018 |
| sd | F | (0.009) | | (0.016) | (0.001) | (0.006) | (0.002) |
| % $\rho - 1$ | | | | 33.2 | 6.4 | 35.7 | 24.7 |
| quarterly frequency | | | | | | | |
| coef | | 0.590 | 2.4 | -0.073 | -0.048 | -0.173 | -0.115 |
| sd | A | (0.011) | | (0.020) | (0.004) | (0.007) | (0.011) |
| % $\rho - 1$ | | | | 17.8 | 11.8 | 42.3 | 28.1 |
| coef | | 0.589 | 2.4 | -0.087 | -0.045 | -0.171 | -0.108 |
| sd | B | (0.011) | | (0.020) | (0.004) | (0.007) | (0.011) |
| % $\rho - 1$ | | | | 21.2 | 11.0 | 41.5 | 26.3 |
| coef | | 0.589 | 2.4 | -0.073 | -0.050 | -0.172 | -0.116 |
| sd | C | (0.011) | | (0.020) | (0.004) | (0.007) | (0.011) |
| % $\rho - 1$ | | | | 17.7 | 12.1 | 41.9 | 28.2 |
| coef | | 0.585 | 2.4 | -0.019 | -0.074 | -0.191 | -0.131 |
| sd | D | (0.016) | | (0.027) | (0.005) | (0.008) | (0.015) |
| % $\rho - 1$ | | | | 4.5 | 17.9 | 46.0 | 31.6 |
| coef | | 0.591 | 2.4 | -0.075 | -0.046 | -0.174 | -0.115 |
| sd | E | (0.011) | | (0.020) | (0.004) | (0.007) | (0.011) |
| % $\rho - 1$ | | | | 18.4 | 11.2 | 42.4 | 28.0 |
| coef | | 0.588 | 2.4 | -0.092 | -0.041 | -0.171 | -0.108 |
| sd | F | (0.011) | | (0.020) | (0.004) | (0.007) | (0.012) |
| % $\rho - 1$ | | | | 22.4 | 10.1 | 41.4 | 26.1 |

Note: Specification A for controls for COVID onset, March to December 2020, and Social Unrest binary variable from October 2019 onwards. Specification B excludes observations after October 2019. Specification C is like A but includes shares of immigrant workers, household heads, married- partners, and placement in four geographical zones (Santiago, North, Center excluding Santiago, and South). Specification D controls for COVID, March 2020 to December 2023, and Social Unrest binary variable from October 2019 onwards. Specification E includes a binary control taking value 1 for March 2020-December 2020, a linear trend, and their interaction. Specification F gets rid of all COVID variables and linear trends for the whole sample. All specifications include fixed effects for bin variables: gender, age group, and education group, as well as monthly seasonal dummies.

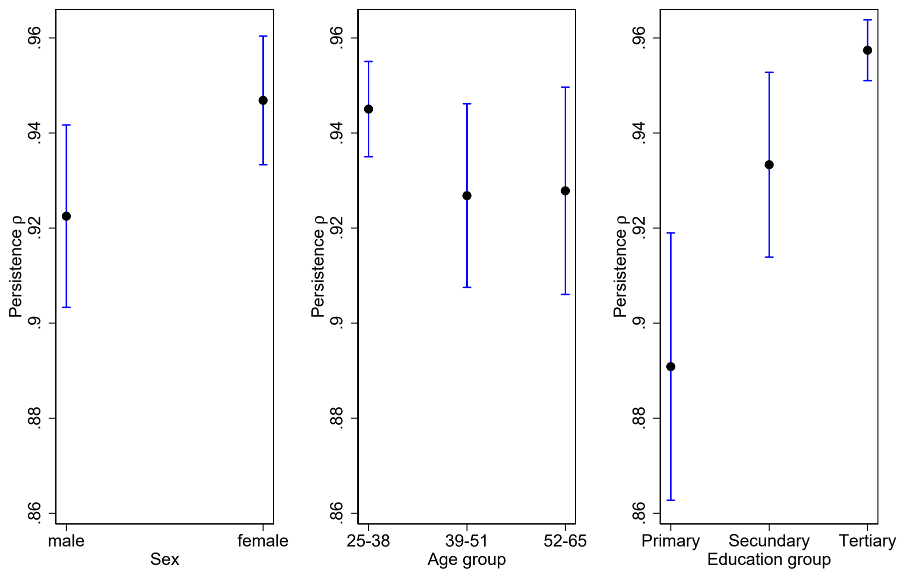
Figure A17: Heterogeneous weekly persistence by groups, DFL weights, flow lags




Note: results are based on a specification of local projections including transition lags and controls for COVID onset, March to December 2020; Social Unrest binary variable from October 2019 onwards, and monthly seasonal dummies.

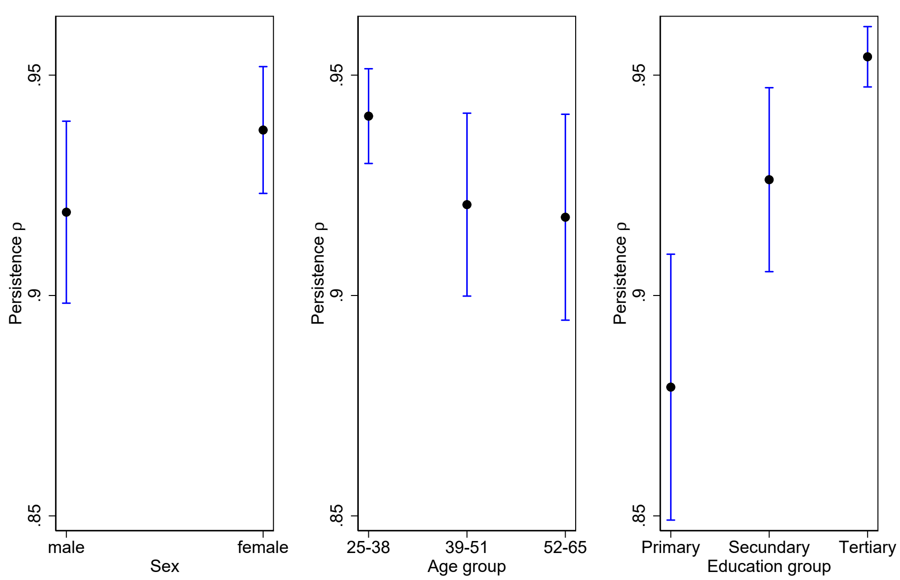
## A.4 Additional heterogeneity results

Figure A18: Heterogeneous weekly persistence by groups, survey weights



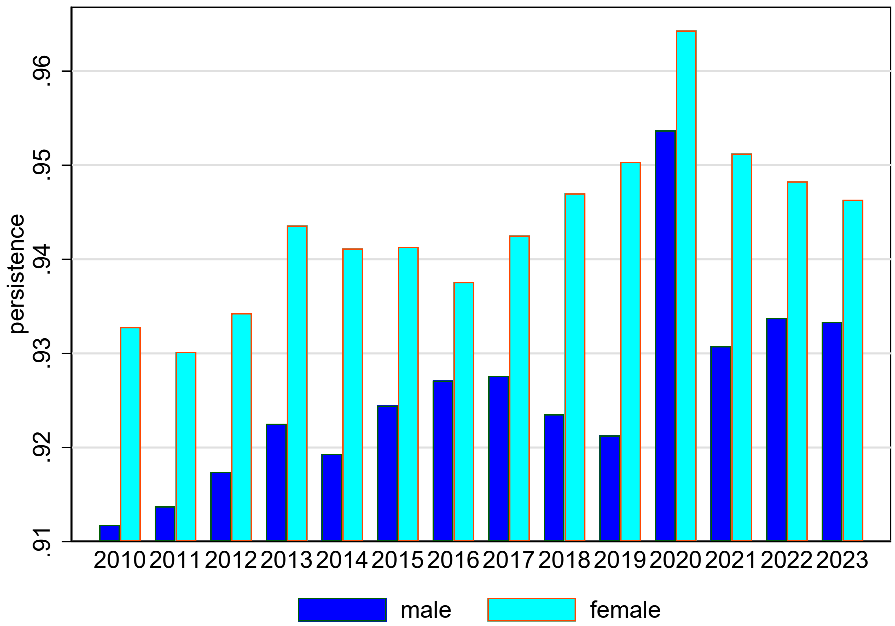

Note: results are based on a specification of local projections without transition lags and controls for COVID onset, March to December 2020; Social Unrest binary variable from October 2019 onwards, and monthly seasonal dummies.

Figure A19: Heterogeneous weekly persistence by groups, survey weights, flow lags



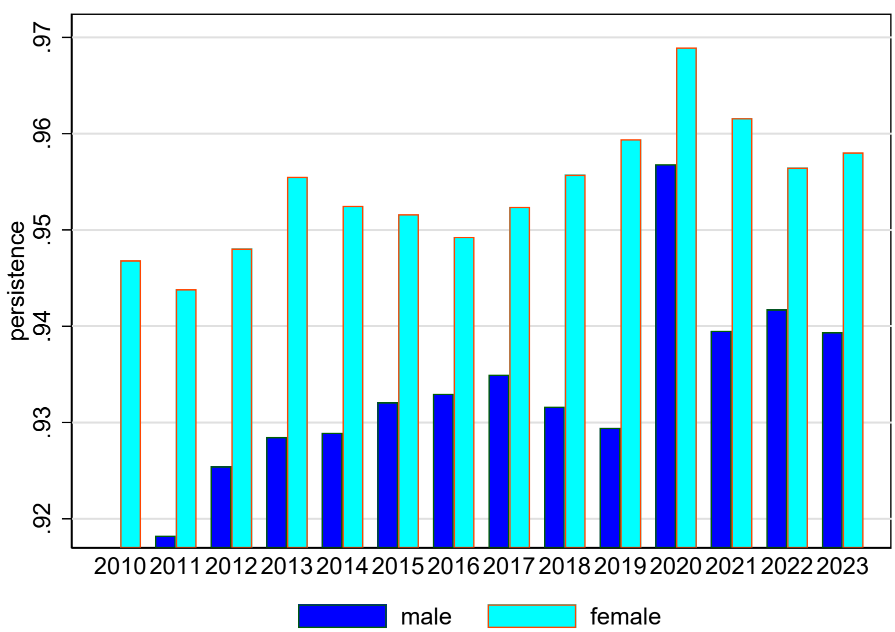

Note: results are based on a specification of local projections including transition lags and controls for COVID onset, March to December 2020; Social Unrest binary variable from October 2019 onwards, and monthly seasonal dummies.

Figure A20: Median weekly persistence over time by gender, DFL weights, flow lags



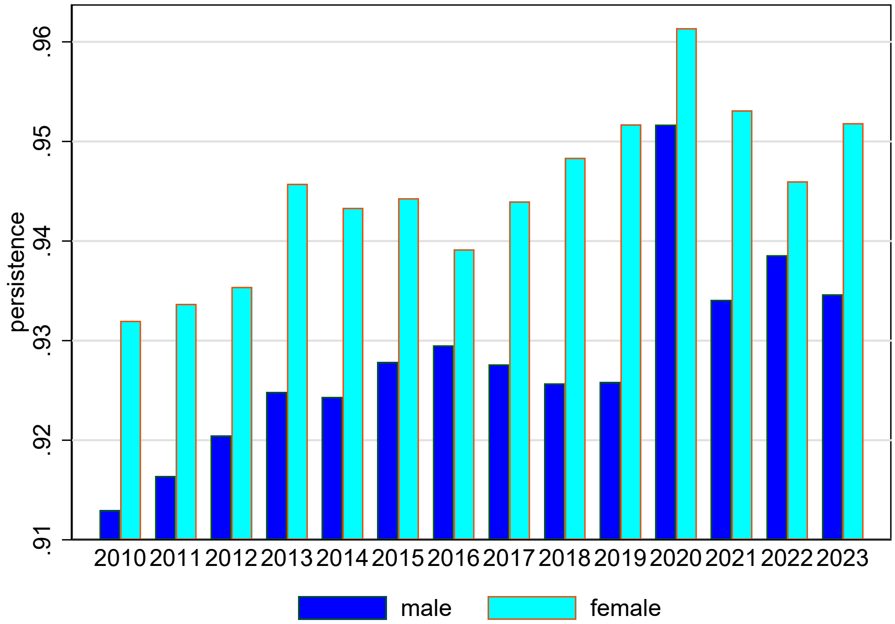

Note: results are based on a specification of local projections including transition lags and controls for COVID onset, March to December 2020; Social Unrest binary variable from October 2019 onwards, and monthly seasonal dummies.

Figure A21: Median weekly persistence over time by gender, survey weights



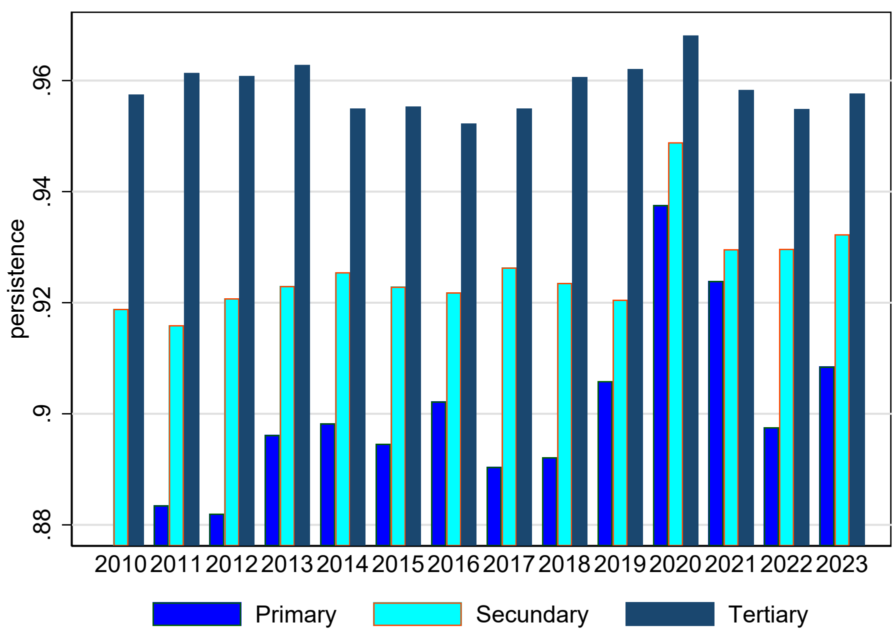

Note: results are based on a specification of local projections without transition lags and controls for COVID onset, March to December 2020; Social Unrest binary variable from October 2019 onwards, and monthly seasonal dummies.

Figure A22: Median weekly persistence over time by gender, survey weights, flow lags



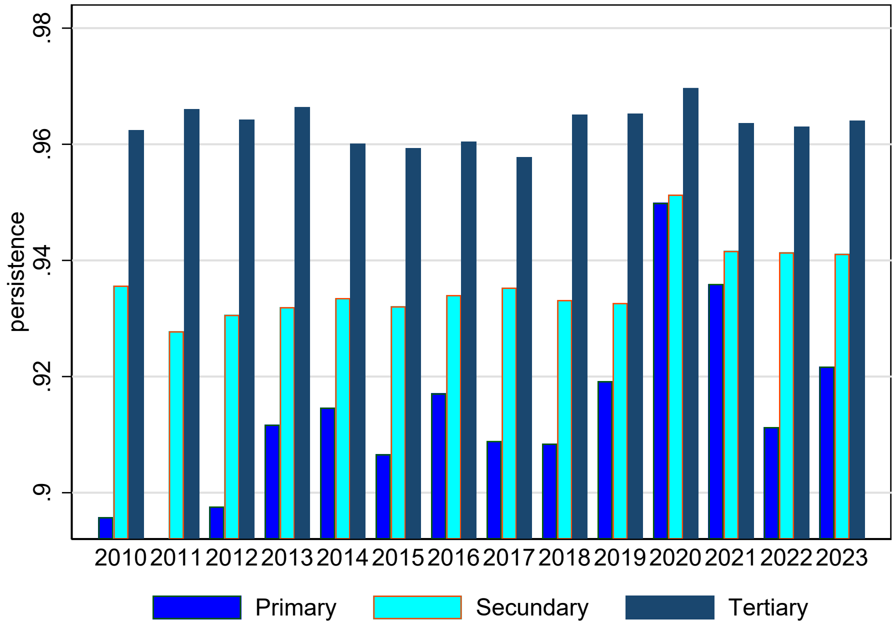

Note: results are based on a specification of local projections including transition lags and controls for COVID onset, March to December 2020; Social Unrest binary variable from October 2019 onwards, and monthly seasonal dummies.

Figure A23: Median weekly persistence over time by college status, DFL weights, flow lags



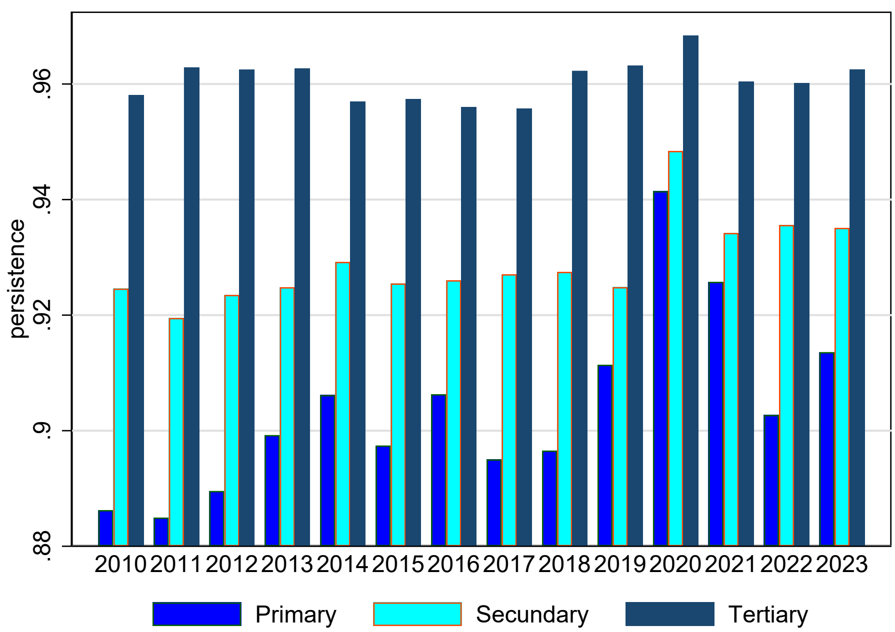

Note: results are based on a specification of local projections including transition lags and controls for COVID onset, March to December 2020; Social Unrest binary variable from October 2019 onwards, and monthly seasonal dummies.

Figure A24: Median weekly persistence over time by college status, survey weights




Note: results are based on a specification of local projections without transition lags and controls for COVID onset, March to December 2020; Social Unrest binary variable from October 2019 onwards, and monthly seasonal dummies.

Figure A25: Median weekly persistence over time by college status, survey weights, flow lags




Note: results are based on a specification of local projections including transition lags and controls for COVID onset, March to December 2020; Social Unrest binary variable from October 2019 onwards, and monthly seasonal dummies.

## A.5 Additional tables

Table A5: Logit models for DFL adjustment

| Dependent variable | Year 2017 |
|---|---|
| covariates | |
| | |
| female | 0.217*** |
| | (0.0608) |
| Chilean | -0.0396 |
| | (0.0392) |
| female X Chilean | -0.0910* |
| | (0.0522) |
| head | 0.146*** |
| | (0.0126) |
| female X head | -0.0968*** |
| | (0.0164) |
| Married | -0.0339 |
| | (0.0293) |
| partner not married | 0.0348 |
| | (0.0313) |
| single | 0.0303 |
| | (0.0314) |
| divorced | 0.0434 |
| | (0.0348) |
| female X married | 0.00343 |
| | (0.0331) |
| female X partner | -0.0341 |
| | (0.0354) |
| female X single | -0.139*** |
| | (0.0342) |
| female X divorced | -0.110*** |
| | (0.0390) |
| Observations | 3,732,932 |

Note: Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$. The estimated logit have a dependent variable that takes value 1 if the observation took place in 2017.
Other controls not shown in the table are categorical variables for age group, detailed 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).