Borradores de ECONOMÍA

Estimating Vacancy Stocks from Aggregated Data on Hires: A Methodology to Study Frictions in the Labor Market.

By: Leonardo Fabio Morales
Eleonora Dávalos
Raquel Zapata

No. 1228
2023

tá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Col

Estimating Vacancy Stocks from Aggregated Data on Hires: A Methodology to Study Frictions in the Labor Market.

Leonardo Fabio Morales	Eleonora Dávalos	Raquel Zapata
lmoralzu@banrep.gov.co	edavalosa@eafit.edu.co	rszapatav@unal.edu.co

Abstract

We develop a methodology that recovers an estimate of the average stock of vacancies using the information on aggregated hires. We show that our prediction of the vacancy stock is unbiased, and it captures well the level and the dynamics of the United States job opening positions reported in the Job Openings and Labor Turnover Survey. We use the methodology to predict vacancies in Colombia for formal and informal salaried workers; together with unemployment, we estimate Beveridge curves and matching functions by occupations, which allows us to study the nature of the efficiency, frictions, and mismatches for different occupations. We find that the formal labor market of technicians is the most inefficient of them all; this inefficiency comes from the mismatch between the abilities of the workers and the requirement of the vacancies. Reducing friction in this occupation will require education and job-oriented training policies. In contrast, the frictions in the market for unskilled workers come from informational lacks. The reductions of friction, in this case, will come from better intermediation and active search policies.

Keywords: Vacancies, labor demand, labor market frictions.
JEL Classification Codes: J60, J63, J23

Estimación de vacantes desde datos agregados de contrataciones: una metodología para estudiar las fricciones del mercado laboral.

Leonardo Fabio Morales	Eleonora Davalos	Raquel Zapata
lmoralzu@banrep.gov.co	edavalosa@eafit.edu.co	rszapatav@unal.edu.co

Resumen

Este trabajo desarrolla una metodología de estimación del stock vacantes a partir de información de contrataciones agregadas. Mostramos que nuestra predicción es consistente en la medida que captura el nivel y la dinámica de las vacantes recolectadas en la Encuesta de Vacantes y Rotación Laboral (JOLTS) en los Estados Unidos. Como una aplicación de la metodología, el trabajo predice las vacantes en Colombia para trabajadores asalariados formales e informales. Posteriormente se estiman curvas de Beveridge y funciones de emparejamiento por ocupaciones, lo que permite estudiar la naturaleza de la eficiencia, las fricciones y los desajustes para los sub-mercados laborales de diferentes ocupaciones. Se encuentra que el mercado laboral formal de técnicos es el más ineficiente de todos; esta ineficiencia proviene del desajuste entre las capacidades de los trabajadores y el requerimiento de las vacantes. Reducir la fricción en esta ocupación requerirá políticas de educación y formación orientadas al trabajo. En cambio, las fricciones en el mercado de trabajadores no calificados provienen de carencias de información. Las reducciones de fricciones, en este caso, vendrán de mejores políticas de intermediación y búsqueda activa.

Keywords: Vacantes, demanda laboral, fricciones.
JEL Classification Codes: J60, J63, J23

1. Introduction

Vacancies play a fundamental role in modern labor macroeconomics, both from a theoretical and empirical point of view. They are crucial in the most salient recent theoretical developments as matching functions. Vacancies play a fundamental role in determining unemployment and the formation of new employer-employee matches. The relationship between job seekers and job openings, better known as the Beveridge curve, is an essential tool for understanding the efficiency of the matching process that generates new hires, as well as the nature of shocks that produce fluctuations in the labor market (Elsby, Michaels, and Ratner, 2015). From the empirical point of view, the measurement of vacancy stocks allows taking to the data these theoretical developments. Furthermore, from the observation of vacancy stocks, we can have a deeper understanding of labor demand dynamics and the nature of frictions and mismatches affecting the efficiency of the hiring process of a particular labor market.

The most natural definition for the concept of vacancy is an unmet labor demand (Abraham, 1983). Therefore, a vacancy might be understood as a job position still waiting to be filled by a suitable worker. A precise measurement of vacancy stocks requires very specialized longitudinal studies at the level of establishments, where the firms' payroll is periodically sized, and they report the stock of job opening positions at a given period. Even though datasets of such complexity have been developed in recent decades, they are only available for very few countries worldwide.

The most reliable data on vacancies in the United States (US) is available from December of 2000 on. This measurement of open job positions is possible with the adoption of the Job Openings and Labor Turnover Survey (JOLTS), a study from the US Bureau of Labor Statistics (BLS). The survey produces monthly estimates of job openings, hires, and separations for a representative sample of all States in the US. This survey study marked a new era in the practice of vacancy measurement. The specific definition of a vacancy in JOLTS is a job position open on the last business day of the reference month, which could start within 30 days. In addition, the employer must actively perform recruitment activities to fill the position with workers from outside the establishment. The survey excludes from the vacancy stock of the establishment internal transfers, promotions, or demotions (US Bureau of Labor Statistics, 2020).

Before the JOLTS in the US, vacancy measurement was mainly based on alternative indirect methods; the most traditional was help-wanted indexes. These indexes are based on the counting of the help-wanted post in newspapers. A similar situation is observed in developing economies; in the absence of proper measurement tools such as JOLTS, the estimation of the vacancy stock in these countries is either based on newspaper help-wanted indexes (Arango, 2013). Also, some studies use analogous methodologies of counting job advertisement posts in job portals (Morales, Ospino, Amaral, 2021).

With the popularization of the internet, help-wanted posting has become very uncommon. Thus, the indexes based on counting these posts need corrections for their validity (Barnichon, 2010). Regarding open positions posted in job portals, it is not easy to understand the representativeness of this subsample of vacancies. A good case of study in this matter is Colombia; since 2015, firms have been required by law to report all open job positions in a public information system. All the information is posted and managed by a public office known as the Public Employment Service (SPE acronym in Spanish). During 2019, 5 years after the implementation of the SPE, the total vacancies posted in the SPE account for less than 20% of the total formal hires observed for that year in administrative records of the social security system (Morales, Ospino, and Amaral; 2021). Therefore, the universe of all vacancies posted online is a small fraction of all vacancies generated that year in the labor market.

In this paper, we propose a methodology that recovers an estimate of the average stock of vacancies, using the information on aggregated hires for this purpose. The methodology is based on the idea that the stock of vacancies can be represented as a forwarded polynomial of hires. In other words, monthly vacancies will be filled in the current and subsequent months, translating vacancies into new hires simultaneously and in future periods. We use the information on total hires per economic sector in the US to validate our estimations. The data come from the public access of JOLTS. This survey reports new open positions and hires aggregated by sectors. We show that our prediction of the vacancy stock captures well the level and the dynamics of the observed job opening positions in JOLTS. Most of the aggregated JOLTS open job positions observed series are contained in the 95% confidence interval of our prediction. The methodology can be applied to other aggregations such as cities, labor market segments, or occupations. In contrast to vacancies, information on hires is widely available from administrative records, or even regular household surveys, for many economies around the globe.

As an application of the methodology, we estimate vacancy stocks in the Colombian labor market. With the information on vacancies and unemployment, we represent Beveridge curves (BC) for the market of formal and informal salaried workers. The estimated BCs for the formal segment of the market hold the expected properties from theory; they describe a stable negative sloped relationship between vacancies and unemployment. In the case of informal markets, the BC shows a U-shaped curve. The relationship is slightly negative at low levels of unemployment. However, at high levels of unemployment, the relationship between unemployment and informal vacancy rates turns out to be positive. We argue that this behavior results from job opening behavior across the economic cycle. In recessions, when unemployment is excessively high, the new labor market demand in the formal sector is weak. Nevertheless, new labor demand in the informal sector can thrive, given that existing or new firms might opt for reducing costs by not complying with regulations.

In the study's remaining sections, we focus on the formal market. We use Colombia’s official household survey to compute unemployment rates by occupation and then represent these BCs by occupation. Comparing the shapes of different Beveridge curves sheds light on the relative efficiency of the matching process of employers and employees across different occupations. In the formal market, occupations with higher skills requirements as managers and professionals have a more efficient matching process than occupations such as technicians, administrative assistants, machine operators, and other professions with tertiary education requirements, but not at the professional level. Furthermore, the formal markets for contractors and service providers are the ones that exhibited greater levels of inefficiency in the matching process. These inefficiencies can be attributable to informational shortages or a mismatch of abilities between workers and jobs. In order to identify the nature of these frictions, we estimate stock and flow matching functions using our prediction of vacancies, which allows testing by occupations if informational shortages or mismatches explain frictions.

Our findings support the hypothesis that for some occupations as directors/managers, professionals/scientists, and technicians, the inefficiencies are explained to a certain degree for the existence of mismatch. In other occupations, such as professionals/scientists and unskilled workers, the explanation is more the existence of deficiency in the information on where the workers and the vacancies are. These diagnostics are crucial for labor market policy designs because they allow for defining which occupations' efficiency gains will come from better intermediation and active search policies. In order to reduce the labor market friction in occupations with a mismatch, more than an active search policy is required. In such a case, educational policies to increase the productivity of the workers and targeted training in the most demanded skills will be more suitable.

A second application of the methodology uses estimated vacancies from administrative records to assess the tightness of the Colombian labor market after the pandemic. We use the methodology described in Domash and Summers (2022) to compare the actual unemployment rate and the firm-side unemployment rate calculated from the vacancy rate and proxy variables for the quit rate. The results show that the demand-side unemployment is lower for the post-pandemic period than the regular unemployment rate, suggesting a tighter labor market from the demand side.

The rest of the paper follows in the following way. Section 2 describes a theoretical framework that establishes a mapping between aggregated hires and the stock of vacancies. Section 3 describes an algorithm to estimate vacancy stock from data on aggregated hires. Section 4 describes all the data sources we use in this paper. Section 5 presents a validation test of our methodology using JOLTS aggregated data. In section 6, we apply our methodology to estimate matching functions and Beveridge curves by different occupations. Section 7 proposes another application for estimating the post-pandemic firm-side unemployment rate. Finally, in section 8, we conclude and offer some policy recommendations.

2. A Theoretical Framework of Aggregated Hires and Vacancy Stock

In this framework, we propose a relationship between vacancies and hires that illustrates that a firm's hiring is the mechanism through which vacancies are filled. This relationship comes from an accounting premise: hires are generated for two reasons, (1) to create new job positions and (2) to replace workers who left. In both previous cases, vacancies need to be filled through hiring, but because there are frictions in the labor markets and the search process is costly, vacancies are not filled simultaneously. A proportion of the vacancies will wait some time to match a worker. Previous literature has explored this idea; for instance, Lazear and Spletzer (2012) decompose hiring into growth and replacement. Morales and Lobo (2020) use a similar distinction for the flow of vacancies, categorizing vacancies into two types: expansion, and replacement vacancies. The methodology we develop in this paper maps hiring to vacancies so that we can take advantage of aggregate information by segments of the labor markets. This consideration is convenient because, in many cases, open-access data sets include hires, our key variable, aggregated by segments of the labor markets: cities, economic sectors, and occupations.

Let us consider firm j in the segment s of the labor market. The firm hires at a given period ($h_{j,s,t}$) are a function of the flow of new vacancies in the current period and in previous periods. In the following equation, the flow of new vacancies generated in firm j, in segment s, at time t is represented by $\underline{v}_{j,s,t}$. Equation (1) represents the idea that hirings are current or previous vacancies being filled at period t; we assume that, on average, vacancies could take up to R periods to be filled.

$$h_{j,s,t} = \phi_0^s \underline{v}_{j,s,t} + \alpha_1^s \underline{v}_{j,s,t-1} + \cdots + \alpha_R^s \underline{v}_{j,s,t-R} \qquad (1)$$

In equation (1) hires are a function of the flow of vacancies instead of the stocks, which is inconvenient because vacancy flows are much more complex to measure than stocks; we would express equation (1) in terms of the stocks rather than flows. In appendix (A), we show a deterministic relationship between new vacancies and the stock of vacancies. At any given period t, the vacancy stock includes the flow of new vacancies generated at that period and part of the flow of new vacancies generated in previous periods that still need to be filled. The flow of new vacancies can be represented as a fraction of the stock of vacancies $\underline{v}_{j,s,t} \approx$

$\underline{\alpha_0^s} \cdot v_{j,s,t}$, and each one of the summands in the left-hand side of equation (1) can be represented as $\phi_\tau^s \underline{v}_{j,s,t} \approx \underline{\phi_\tau^s \alpha_\tau^s} \cdot v_{\tau,s,t}$. Therefore, the following equation is a representation of hires $(h_{j,s,t})$ of average firm j in segment s, as a function of the stock of vacancies $(v_{j,s,t})$ in the current and previous periods:

$$h_{j,s,t} = \alpha_0^s v_{j,s,t} + \alpha_1^s v_{j,s,t-1} + \cdots + \alpha_R^s v_{j,s,t-R} \qquad (2)$$

$$\sum_{j \in s} h_{j,s,t} = \alpha_0^s \sum_{j \in s} v_{j,s,t} + \alpha_1^s \sum_{j \in s} v_{j,s,t-1} + \cdots + \alpha_R^s \sum_{j \in s} v_{j,s,t-R}$$

$$H_{s,t} = \alpha_0^s V_{s,t} + \alpha_1^s V_{s,t-1} + \cdots + \alpha_R^s V_{s,t-R} \quad (3)$$

In equation (3) $H_{s,t}$ and $V_{s,t}$ represents hires and vacancy stocks aggregated by segments, respectively. Equation (2) can be inverted to find expressions for $v_{j,s,t}$, in doing so, the following system of equations is generated:

$$\frac{1}{R} v_{j,s,t} = \frac{\left[h_{j,s,t} - \left(\alpha_1^s v_{j,s,t-1} + \cdots + \alpha_R^s v_{j,s,t-R}\right)\right]}{\alpha_0^s R} \cong \beta_0^s h_{j,s,t} \qquad (4)$$

$$\frac{1}{R} v_{j,s,t} = \frac{\left[h_{j,s,t+1} - \left(\alpha_1^s v_{j,s,t} + \cdots + \alpha_R^s v_{j,s,t-R+1}\right)\right]}{\alpha_1^s R} \cong \beta_1^s h_{j,s,t+1} \quad (5)$$

$$\vdots$$

$$\frac{1}{R} v_{j,s,t} = \frac{\left[h_{j,s,t+R} - \left(\alpha_1^s v_{j,s,t+R} + \cdots + \alpha_R^s v_{j,s,t}\right)\right]}{\alpha_R^s R} \cong \beta_R^s h_{j,s,t+R} \quad (6)$$

Equations (4) to (6) come from solving for $v_{j,s,t}$ in current and forwarded versions of equation (2). These equations represent the idea that a fraction of the stock of vacancies of the average firm in segment s is filled in current and subsequent periods. The summation from equation (4) to equation (6) generates the following expressions:

$$v_{j,s,t} \cong \beta_0^s h_{j,s,t} + \beta_1^s h_{j,s,t+1} + \cdots + \beta_R^s h_{j,s,t+R} \qquad (7)$$

$$\sum_{j \in s} v_{j,s,t} = \sum_{j \in s} \beta_0^s h_{j,s,t} + \sum_{j \in s} \beta_1^s h_{j,s,t+1} + \cdots + \sum_{j \in s} \beta_R^s h_{j,s,t+R}$$

$$V_{s,t} = \beta_0^s H_{s,t} + \beta_1^s H_{s,t+1} + \cdots + \beta_R^s H_{s,t+R} \qquad (8)$$

The last equation aggregates the stock of vacancies and hires over the labor market segment. Equation (8) expresses that the aggregate stock of vacancies at the current period will be partially filled in the simultaneous and subsequent periods.

3. An Algorithm to Estimate Vacancy Stock from Aggregated Hires Data

We propose estimating aggregated vacancies using hires based on equation (8) and equation (3). In household surveys or administrative records, aggregated data on aggregated hires are available for many countries, but data on aggregated vacancies are scarce. In the absence of information on vacancies, the variable $V_{s,t}$ is unobserved in the equation (8); nevertheless, we can write this equation as:

$$H_{s,t} = \left[V_{s,t} + \beta_1^s H_{s,t+1} + \cdots + \beta_R^s H_{s,t+R}\right] * \frac{1}{\beta_0^s} \quad (9)$$

Even though in the previous $V_{s,t}$ it is still unobservable, an estimable version of equation (9) might be written as:

$$H_{s,t} = \delta_{s,t} + \beta_1^s H_{s,t+1} + \cdots + \beta_R^s H_{s,t+R} + \varepsilon_{s,t} \quad (10)$$

$$where\ \delta_{s,t} = \left[\delta_s + \delta_{month,year} + \delta_{quarter} + \delta_{s,year} + \delta_s * Trend + \delta_s * Trend^2\right]$$

In equation (10), the unobservable variable $V_{s,t}$ is estimated by residual using the coefficient for the specific segment and time-varying intercept of the equation. In equation (10) estimation, we allow all coefficients to vary by occupation and time. Regarding the latter, we estimate equations using moving windows of several years; this allows the coefficients to vary in each window.

A remaining question is how to determine the polynomial length in equation (10); for this purpose, we use equation an estimable version of equation (3). We estimate a version of equation (3) from a set of possible specifications, in which we replace the unobserved variables $V_{s,t}$, by its estimated versions $\hat{\delta}_{s,t}$. This second estimated equation can be represented as:

$$H_{s,t} = \alpha_0^s \hat{\delta}_{s,t} + \alpha_1^s \hat{\delta}_{s,t-1} + \cdots + \alpha_R^s \hat{\delta}_{s,t-R} + u_{s,t} \quad (11)$$

We estimate a set of regressions (11) for a set of different specifications of equation (10); finally, we choose the preferred specification using the Bayesian information criterion (BIC = -2*log-likelihood+Klog(n)). We estimate equations and vacancy stocks predictions for a set of specifications with different lengths of the polynomial (R), and finally, we choose the specification with the smallest BIC. Finally, we choose estimation $\hat{\delta}_{s,t}$ as the best estimator for $V_{s,t}$ from that specification. In appendix C, we illustrate an equation (10) estimation for the optimal polynomial length in a given window.

4. Data

We use US aggregated data from JOLTS, which is openly available from the US Bureau of Labor Statistics to validate the methodology. The survey uses a sample of nearly 20,700 establishments, out of a population of 9.4 million, from public and private sectors. The sample is representative of all non-agricultural economic sectors in the US and District of Columbia. JOLTS is a rotating panel; new establishments are incorporated into the sample every month and are followed for 24 months. After this period, the establishments exit the sample and will not return during the subsequent three years. Therefore, each month establishments enter and exit the sample continuously. We will use a panel of 18 economic sectors: Mining and Logging, Construction, Durable goods manufacturing, Nondurable goods manufacturing, Wholesale trade, Retail Trade, Transportation, Warehousing and utilities, Information, Finance and insurance, Real estate and rental and leasing, Professional and business services, Educational services, Health care, and social assistance, Arts, entertainment, and recreation, Accommodation and food services, Federal, State and local, and finally Other services. We can observe aggregated open job positions and hires for each of these economic sectors. In section 5, we show that we can obtain unbiased estimates of the observed stock of vacancies using only information on hires.

As an application of our methodology, we use the information on hires computed from the official Colombian household survey GEIH (for its acronym in Spanish). The GEIH is a standard household survey; it is the official source of labor market statistics. As with many other household surveys, the survey asks employees how long they have had their current position. This question is the basis of the computation of total hires in the labor market segment; we define as hires all matched employer-employee for which the employee reports a job tenure of one month or less. In this application, we aggregate hires into eight different occupations: (1) managers, directors, and CEOs, (2) professionals and scientists, (3) technicians, (4) administrative assistants, (5) service providers and sellers, (6) contractors, (7) machine operators, (8) unskilled occupations. Table 1 shows summary statistics of the labor market variables we use in our application for the Colombian labor market, and Appendix D shows the same statistics disaggregated by occupation.

Table 1: Descriptive Statistics

Variables	Obs.	Mean	Median	Std.	Min	Max
Employment	1,248	1,362,240	1,034,120	1,172,185	211,469	4,697,535
Formal salaried workers	1,248	557,530	497,087	323,169	72,680	1,487,507
Unemployed	1,248	171,585	119,890	168,296	11,624	716,784
Short unemployment	1,248	56,746	34,957	59,959	1,326	373,823
Hires	1,248	30,638	25,710	22,932	425	126,765

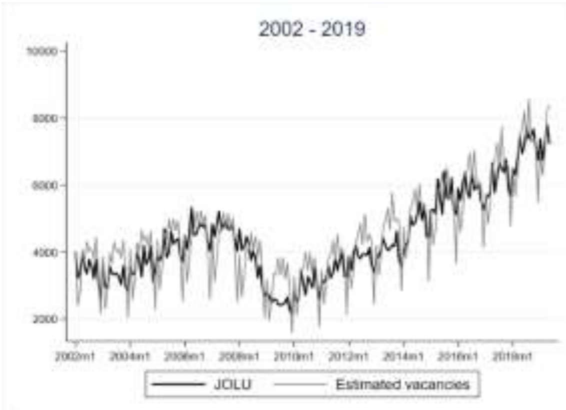
Notes: This table summarizes the descriptive statistics of the variables used in the methodology application. The sample corresponds to a panel of 8 occupations with monthly data between January 2007 and December 2019.

5. Validation of the Methodology Using JOLTS

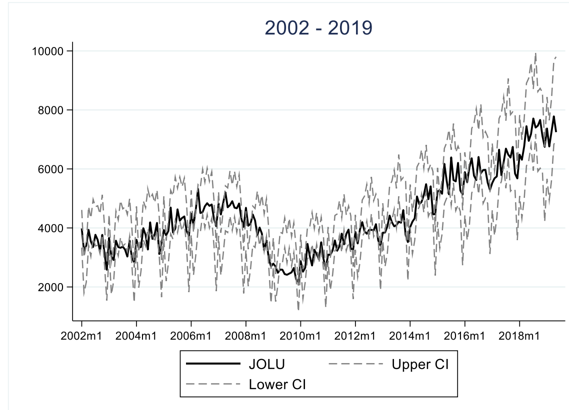
We apply the methodology presented in section 3 using open-access JOLTS data. We use data on aggregated hires by economic sectors. The estimation procedure develops as follows. First, to allow that coefficients vary over time, we estimate equation (10) using four-year windows; we estimate specifications from one to five polynomial lags. We use the BIC information criteria to choose the preferred specification in each window estimation. Since we have overlaps across different windows' estimations, we also use the prediction with the lowest BIC. Finally, we compute standard errors of the predictions by bootstrap methods; we generate random samples from the aggregated hiring data, and we use 250 replications.[1]

[1] In Appendix B, we show that estimations are similar when we choose different lengths for the estimation windows.

Graph 1: JOLTS vacancy stock and point estimated prediction.



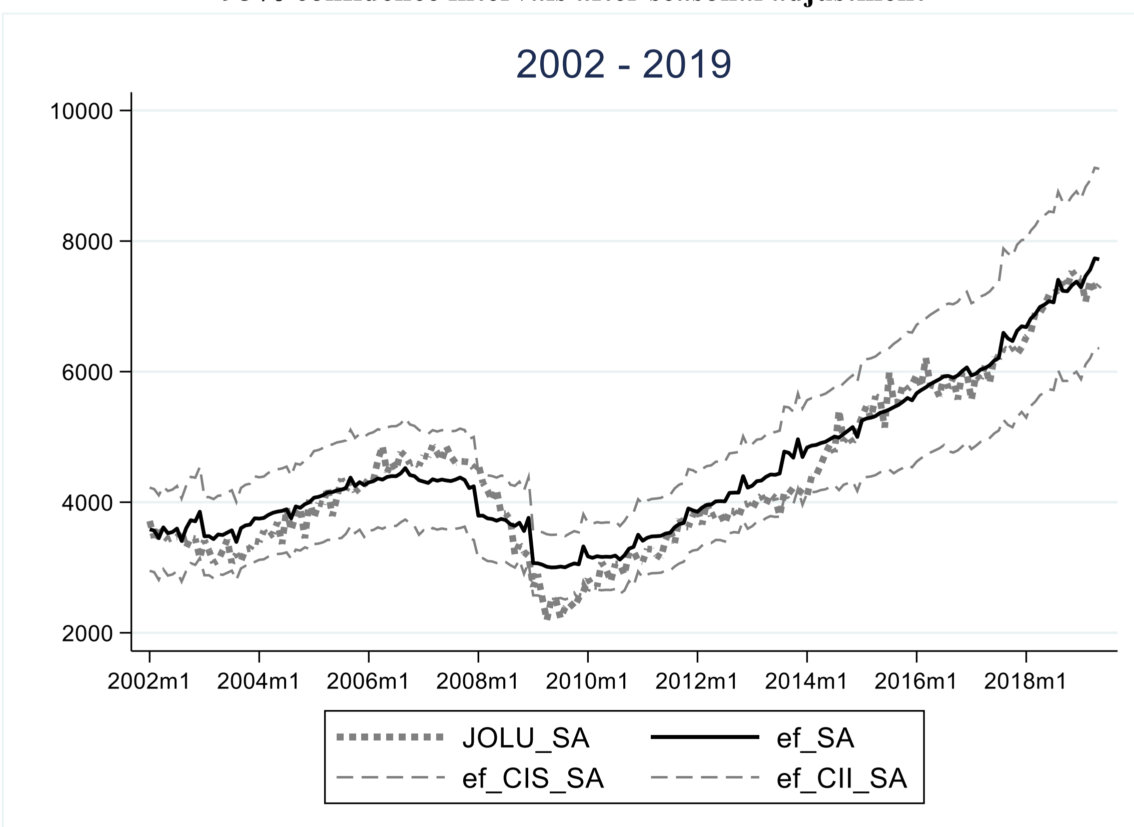
Graph 2: Prediction's Confidence Intervals.



Notes: Graph 1 compares the aggregated vacancies from JOLTS and the predicted series of the stock of vacancies, computed using the methodology presented in section 3.

Notes: Graph 2 shows the aggregated vacancies from JOLTS and the confidence intervals constructed at a 95% confidence level. The bootstrap was parametrically performed with 250 replications.

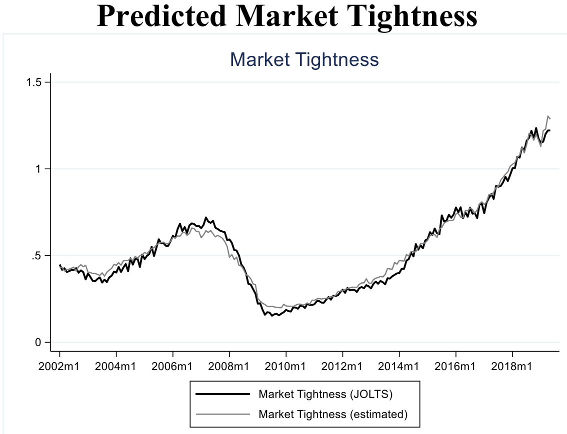
Graph 3: Comparison of JOLTS vacancy stock and point estimated prediction with 95% confidence intervals after seasonal adjustment



Notes: Graph 3 shows the seasonally adjusted series of the observed and predicted stock of vacancies and the confidence intervals constructed at a 95% confidence level.

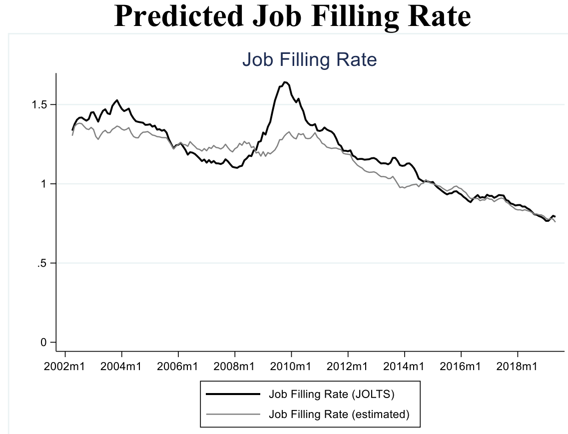
Graphs 1, 2, and 3 show the results of our methodology. In Graph 1, we show an aggregated prediction of the stock of vacancies adding up the stocks for each sector; we compare this prediction with the aggregated vacancies from JOLTS, in both cases, without any seasonal adjustment. In general, the estimation does a good job of predicting the level and dynamics of the stock vacancies. Graph 2 shows that in most periods, the observed vacancies stock is contained in the 95% confidence interval of the prediction. The observed and the predicted series of vacancies show a noticeable seasonal fluctuation, especially the predicted one. Graph 3 shows the observed and predicted vacancies after a standard X-12-ARIMA seasonal adjustment procedure. After seasonal adjustment, we confirm that the prediction is an unbiased estimated estimator of vacancies stock; again, the seasonally adjusted series of the JOLTS vacancies is almost always inside the 95% confidence interval of the prediction with seasonal adjustment. Finally, in Graphs 4 and 5, we show the prediction of additional labor slackness indexes; we correct for seasonal adjustment in both cases. In Graph 4, we show predicted tightness (Vacancies/unemployed). The prediction performance is quite good in capturing the series' dynamics and level. In the case of the Job Filling rate, the methodology performs worse than in the previous case; nevertheless, it captures the level of the series and the general trend.

Graph 4: Comparison of JOLTS & Predicted Market Tightness



Notes: Graph 4 shows the seasonally adjusted series of the observed and predicted ratio of Vacancies/Unemployed.

Graph 5: Comparison of JOLTS & Predicted Job Filling Rate



Notes: Graph 5 shows the seasonally adjusted series of the observed and predicted ratio of Hires/Vacancies.

6. Exploring the inefficiency of Employer-Employee match using Beveridge Curve by Occupation.

This section illustrates the benefits of our methodology in an application that uses our estimation of vacancies for different submarkets, defined as a set of occupations in the labor market. For this purpose, we use conventional information from the Colombian household survey, the GEIH, which was described in section 4. Using the stock of vacancies, we estimate Beveridge curves and matching functions; these tools allow us to describe the nature of labor market frictions and mismatches for different occupations in the Colombian market. Some theoretical background is needed to tackle these topics; we make a succinct revision of the main concepts in subsection 6.1.

6.1 Search and matching frictions:

The Matching function

The unemployment equilibrium model is the modern canonical framework for studying labor market frictions and mismatch (Pissarides, 2000; Blanchard & Diamond, 1989). The fundamental building block of these models is the aggregated matching function. Successful matches between employers and employees are represented as a function of firms' vacancies and the number of unemployed workers. The matching function is a simple characterization of the hires (new matches) in an environment with imperfect information and market frictions (Anderson & Burges, 2000). It is usually assumed as a homogenous function, which can be represented as in the following equation:

$$H = m(U, V) \qquad (12)$$

where H represents the hires, U and V represent the stock of unemployed and vacancies, respectively.

Recent literature has remarked that submarkets are separated by location and occupation (Andrews, Bradley, Stott, and Upward, 2013). Vacancies are generated in each submarket, and workers search for a job in specific locations or occupations. Models that allow free entry into submarkets, where firms and workers can post vacancies and search for jobs in the submarkets they choose, are directed search models (Moen, 1997; Acemoglu & Shimer, 1999). In the following subsection, we will use this concept of a submarket to estimate matching functions and Beveridge curves of submarkets, where a submarket is defined as a specific occupation.

The Beveridge curve

From the matching function (12), and the fact that it exhibits constant returns to scale, surges a negative relationship between vacancies and unemployment. Before developing search and matching models, William (Beveridge, 1944) had established a negative relationship between vacancies and unemployment in statistical terms; for this reason, this relationship is known as the Beveridge Curve (BC). In the canonical model of equilibrium unemployment, the BC is derived in equilibrium from the law of motion of unemployment; for the sake of simplicity, we adopt a more straightforward derivation proposed by Petrongolo and Pissarides (2001). We denote the total labor force as L, and the occupied population as N; therefore, the vacancy and unemployment rates would be $v = V/N$ and $u = U/L$, respectively. In steady-state, unemployment levels are invariant; therefore, the hiring rate, $h = H/N$, is equivalent to the separations rate $s = S/N$, where H and S stand for hires and separations, respectively. Therefore using the homogeneity of (12), in steady-state, we have the following equation:

$$s = \frac{S}{N} = m\left(\frac{U}{L}\frac{L}{N}, \frac{V}{N}\right) = m\left(\frac{u}{1-u}, v\right) \qquad (13)$$

For a given positive rate of separations, equation (13) implies a negative relationship between vacancies and unemployment in the steady state. The shape and position of the BC depend upon the matching technology, information summarized by the matching function. There is a straightforward way to assess the efficiency level in a labor market: markets with fewer frictions and mismatch unambiguously would present a BC closer to the origin in a space vacancy rate-unemployment rate. For a specific unemployment rate closer to the origin, BCs show that the equilibrium vacancy rate is low. Therefore, the matching process in a labor market that exhibits closer to the origin BC means that vacancies are filled faster and more efficiently.

Stock and Flow matching

Extensions to the more simplistic matching function in equation (12) offer alternative explanations for the existence of mismatch in the labor market; stock and flow matching is one of these theories. The matching pattern associated with the standard matching function in equation (12) is random; matching formation is random between one side of the market and the other; therefore, successful matches are a function of the stocks of unemployment and vacancies. Frictions in this setting are only the result of incomplete information on the location of the jobs or workers (Sasaki, 2008). In the case of stock and flow matching, agents search for a short period; after this matching round, unmatched vacancies and workers are not matched for each other. Agents on both sides of the market are unmatched because of the lack of suitable employers or vacancies of particular types (Andrews et al., 2013). Therefore, in subsequent matching rounds, the unmatched agents from previous rounds, which belong to the stock of one side of the market, will most likely match the other side's inflow. The literature on stock and flow has expanded recently; some remarkable studies on the topic are the following: Coles (1994), Coles and Smith (1998), Shimer (2007), and Ebrahimy and Shimer (2010).

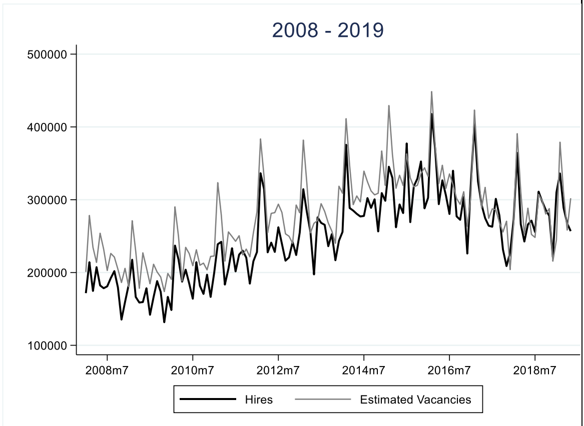
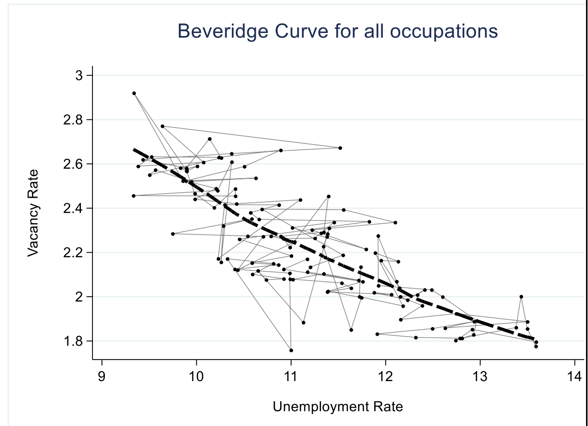
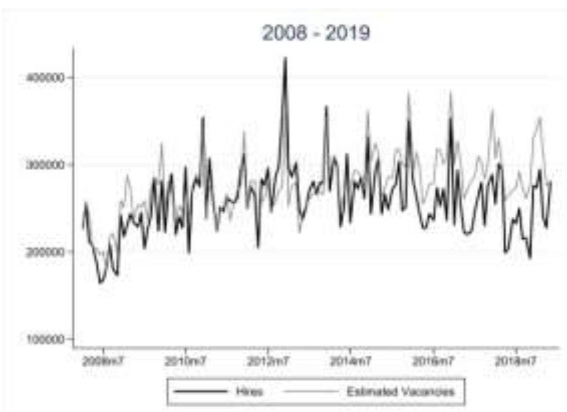
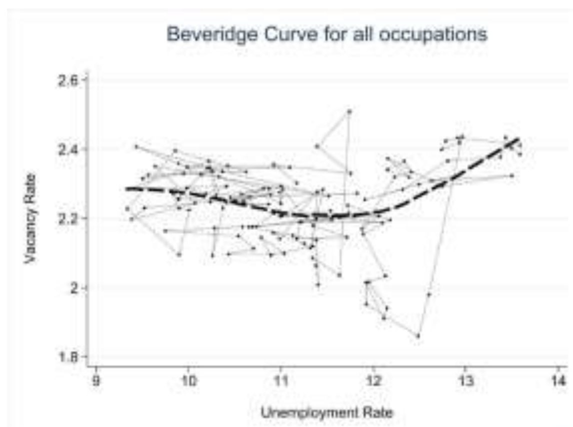
The main argument of stock and flow matching is that the unmatching traders on each side will keep looking because there are no suitable partners on the other side of the market's stocks. Therefore, traders will wait for the next rounds of the search process to match the other side of the market's inflow. The matching pattern has important implications for the policy recommendations for reducing frictions and equilibrium unemployment; this policy implication will differ for random and stock and flow matching patterns. On the one hand, if empirical models support random matching, friction's main explanation is a lack of information. In this latter case, the best policy is enhancing the functional capability of the intermediary institutions in the labor market, such as public or private employment agencies, which are the leading players in executing active labor market policies (Sasaki, 2008).

On the other hand, if empirical models support stock and flow models, some mismatches imply that vacancies and unemployed cannot find suitable partners to match in the stocks of each side of the market. Therefore, policies for training and enhancing workers' skills would be more appealing. Other strategies have been suggested in the literature as the implementation of subsidies for the creation of new jobs in a way that the cost of training on the job would partially be covered; another one is entrepreneurship loans (Dmitrijeva, J. & Hazans, 2005; De Toledo, Núñez, & Usabiaga; 2008).

6.2 Beveridge curves

We apply the methodology presented in section 3, using the information on aggregated hires from the Colombian formal labor market; we measure the hires for eight different occupations using the official Colombian household survey, as described in the data section. Our methodology allows computing vacancies stocks for different labor market segments (submarkets). In this sub-section, we present results for formal and informal salaried workers, defining a formal job as those offered by private companies or the government and for which payroll taxes are paid. Graph 6 shows the aggregation of the stock of vacancies for all occupations for the formal salaried workers; the level of hires for the period 2008-2019 was 280k on average. The estimated stock of vacancies was 310K. In Graph 7, we present an estimation of the aggregate Beveridge curve for this market segment; the BC has the expected properties from the theoretical models of equilibrium unemployment; it depicts a stable, negative sloped relationship between the vacancy rate and the unemployment rate.

Graph 8 shows the aggregation of the stock of vacancies for all occupations for the informal salaried workers; the level of hires for the period 2008-2019 was 257k on average. The estimated stock of vacancies was 270K. Graph 9 estimates the aggregate Beveridge curve for this market segment. To the best of our knowledge, this is the first depiction of an informal BC in the literature. The informal BC shows a U-shaped curve; the relationship is negative at low levels of unemployment, but at high levels of unemployment, it changes slope, and the relationship between unemployment and vacancy rates turns out to be positive. This pattern results from job opening behavior across the economic cycle. In recessions, when unemployment is excessively high, the new labor market demand in the formal sector is weak; few new positions open in a period because job creation is pro-cyclical (Florez et al., 2020; Morales & Medina, 2019). Nevertheless, in a recession, new labor demand in the informal sector can thrive, given that existing or new firms might opt for reducing costs by not complying with regulations.

Graph 6: Vacancy Stock and Hires Formal Salaried Workers	Graph 7: Beveridge curve Formal Salaried Worker Estimated Vacancies
	
Notes: Graph 6 compares the aggregated hires for the formal salaried workers calculated from GEIH and the estimated stock of vacancies computed using the methodology in section 3.	**Notes:** Graph 7 is based on the estimation of equation (13). The thick dotted line shows the estimated Beveridge curve for the segment of the formal salaried workers.
Graph 8: Vacancy Stock and Hires Informal Salaried Workers	**Graph 9: Beveridge curve Informal Salaried Worker Estimated Vacancies**
	
Notes: Graph 8 compares the aggregated hires for the informal salaried workers calculated from GEIH and the estimated stock of vacancies computed using the methodology in section 3.	**Notes:** Graph 9 is based on the estimation of equation (13). The thick dotted line shows the estimated Beveridge curve for the segment of the informal salaried workers.

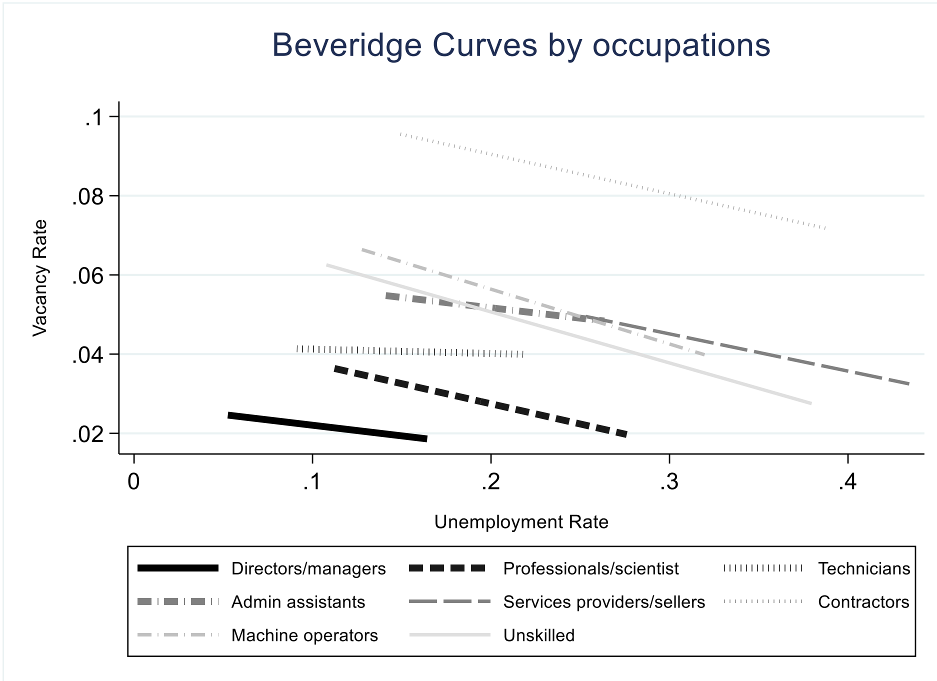
6.2.1. Beveridge curves by occupations

Using information from GEIH, we can compute aggregate hires and unemployment levels by occupation in the formal market; in the case of the latter, the survey asks the responders about the occupation in which they are searching for a job. Our methodology allows computing vacancies for all occupation groups; therefore, we can estimate occupation-specific BC. This exercise applies the theories of directed search; as mentioned in the previous section, recent developments in unemployment equilibrium models remark on the role of submarkets separated by occupations or locations (Andrews et al., 2013). Matching can be heterogeneous across occupations; therefore, the underlying BCs must be heterogeneous; the representation of these functions would shed light on the heterogeneity of frictions and mismatches across occupations. For the sake of brevity from now on, we focus on the formal labor market; nevertheless, the estimation for the informal labor market is presented in Appendix F.

In Graph 10, we present linear estimations of the BC for each occupation. Equilibrium unemployment models indicate that closer to the origin BC is associated with more efficient employer-employee matching. The vacancy rate for a given unemployment level is lower, indicating that the filling vacancies process is more efficient than in the opposite case. In the formal market, occupations with higher skills requirements as managers and professionals have a more efficient matching process than occupations such as technicians, administrative assistants, machine operators, and other professions with tertiary education requirements, but not at the professional level. Furthermore, the formal markets for contractors and service providers are the ones that exhibited more significant levels of inefficiency in the matching process.[2]

[2] In the case of the informal market, as shown in appendix F, we find that managers and professionals are the most efficient informal labor markets. In contrast, unskilled workers and contractors are the most inefficient ones.

Graph 10: Beveridge curve by Occupations with Formal Salaried Worker Vacancies



Notes: Graph 10 shows the linear estimation of the Beveridge curve for each occupation. In the formal market, occupations with higher skills requirements as managers and professionals, have a more efficient matching process, while contractors and service providers exhibit greater levels of inefficiency.

6.3. Explaining the nature of frictions from the matching pattern in the formal labor market

The evidence in the previous section suggests that employer-employee matching is more efficient in some formal occupations than in others. For some occupations as directors, managers, and professionals, the Beveridge curve is closer to the origin than in others as contractors, machine operators, and service providers. These inefficiencies can be explained as informational frictions or structural mismatches; as explained in subsection 6.1, the matching patterns for the former would be random, while the second would be a stock and flow pattern. The literature on estimating matching functions has suggested a clear distinction between flows and stocks. For instance, in the stock and flow hypothesis, non-matched workers from previous periods will not match the stock of vacancies but will match the inflow of new vacancies. Therefore, the identification of matching patterns might be formulated in terms of the specification of the matching function. If only the Stocks of unemployment and vacancies are relevant for forming new matches, then the matching pattern must be random. If inflows of unemployed and vacancies are essential, then stock and flow patterns must occur.

6.3.1. Matching Function and the patterns of matching

To test if there is a pattern of random matching, stock-and-flow, or a combination of the two in each submarket, we will study an augmented stock and flow matching function in which stocks and flows are allowed to play a role in the formation of new matches. We follow the mainstream of the literature and assume that the matching function is a Cobb-Douglas function. Therefore, an augmented matching function can be represented by the following equation:

$$H_{s,t} = \mu_{s,t} * {V_{s,t}}^{\alpha_s^V} * {v_{s,t}}^{\gamma_s^V} * {D_{s,t}}^{\alpha_s^D} * {d_{s,t}}^{\gamma_s^D} * u_{s,t} \qquad (14)$$

$$\ln(H_{s,t}) = \mu_{s,t} + \alpha_s^V \ln(V_{s,t}) + \gamma_s^V \ln(v_{s,t}) + \alpha_s^D \ln(D_{s,t}) + \gamma_s^D \ln(d_{s,t}) + u_{s,t} \quad (15)$$

Applying logarithmic transformation to equation (14), we obtain a linear in-parameters equation (15), in which coefficients are elasticities. In equation (15), $H_{s,t}$ represents the hires in segment s at time t; $V_{s,t}$ and $D_{s,t}$ stands for stocks of vacancies and unemployed in period t, respectively. Finally, $v_{s,t}$ and $d_{s,t}$ stand for flows of new vacancies and new unemployed of segment s in period t. From GEIH, we can directly measure the stock and flow of unemployed, and from the methodology presented in section 3, we can estimate the stock of vacancies. We do not observe the flow of vacancies directly; nevertheless, we can express it as follows:

$$V_{s,t} = V_{s,t-1} - \beta_0^s H_{s,t-1} + v_{s,t} \quad (16)$$

$$v_{s,t} = V_{s,t} - V_{s,t-1} + \beta_0^s H_{s,t-1} \quad (17)$$

Expression (16) describes the stocks of vacancies in the current period as the stock of the previous period minus the fraction of hires in that period that filled vacancies simultaneously; in other words, the stock at the end of the period, plus the new inflow of vacancies in period

t. In equation (17) we obtain an expression for the flow of vacancies $v_{s,t}$ from equation (16). Replacing (17) into (14), we can rewrite the stock and flow matching functions as:

$$\ln(H_{s,t}) = \mu_{s,t} + \gamma_s^V \ln(V_{s,t}) + (\alpha_s^V - \gamma_s^V)\ln(V_{s,t-1}) + \beta_0^s(\gamma_s^V - \alpha_s^V)\ln(H_{s,t-1}) \\ + \alpha_s^D \ln(D_{s,t}) + \gamma_s^D \ln(d_{s,t}) + u_{s,t} \quad (18)$$

From the reduced form estimation (18), we can recover the parameters of equation (14). Specifically, we can estimate parameters: $\widehat{\gamma_s^V}, (\widehat{\alpha_s^V - \gamma_s^V}), \beta_0^s(\widehat{\gamma_s^V - \alpha_s^V}), \widehat{\alpha_s^D}$ and $\widehat{\gamma_s^D}$; we can solve for $\widehat{\alpha_s^D}$, as: $\widehat{\alpha_s^V} = (\widehat{\alpha_s^V - \gamma_s^V}) + \widehat{\gamma_s^V}$. In this way, we can have estimates of all parameters of equation (14): $\gamma_s^V, \alpha_s^V, \gamma_s^D,$ and α_s^D. In order to simplify the interpretation of the coefficients, we use logarithmic transformations of all dependent and independent variables in equation (18). In addition, we include occupation and period fixed effects and heterogeneous linear and quadratic trends by different occupations to control for any endogeneity issue.

6.3.2. Augmented Matching Function Estimation

In Graph 11, we present the plot of coefficients for $\gamma_s^V, \alpha_s^V, \gamma_s^D$ and α_s^D for each occupation, and the results obtained from the estimation of the equation (18) are presented in Appendix E. We obtain the following insights from the estimation of our augmented matching function. From the estimated coefficient $\widehat{\gamma_s^V}$, we find that for all occupations, except for unskilled workers, the coefficient associated with the inflow of vacancies is statistically significant; magnitudes are sizeable for contractors, service providers, machine operators, and administrative assistants. In these occupations, the flow of vacancies is significant in determining new matches; this is a sign of mismatch from the side of vacancies; unmatched unemployed need to wait for additional searching rounds to match with the inflow of new vacancies. Therefore, in the stock of unemployed, there are no suitable or available workers to fill the unmatched vacancies from previous rounds, either because they do not fulfill the requirements of the vacancies stock or because they prefer to wait for the arrival of better quality vacancies.

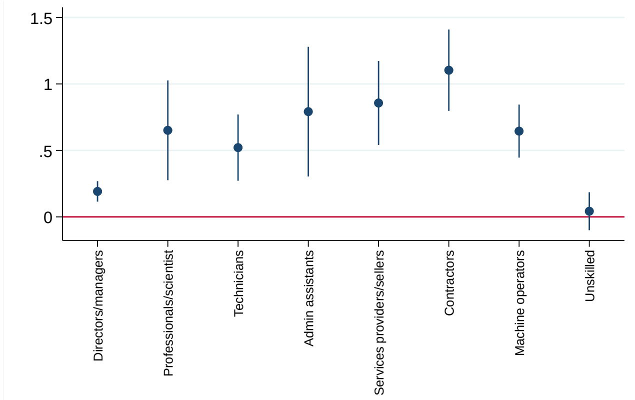
We find evidence of random matching for some occupations from the vacancies' side. In other words, we find a positive and significant α_s^V coefficient in the occupations of directors and managers, professionals/scientists, technicians, and contractors. In these occupations, the stock of vacancies is significantly correlated with hires. This latter correlation reveals that lack of information explains part of the inefficiency in the labor market for these professions on where the vacancies are located.

Regarding unemployment, we find that in the occupations of directors/managers, professionals/scientists, and technicians, the coefficient associated with the inflow of the unemployed population ($\widehat{\gamma_s^D}$) is positive and statistically significant. In these occupations, the flow of unemployed is important in determining new matches; unmatched vacancies from previous rounds must wait for subsequent search rounds to be filled with the new inflow of the unemployed population. The reason for this is that the unemployed workers in the stock for these occupations were not suitable to fill the vacancies in the stock; the new inflow of unemployed is required to generate more matches. As before, this finding is interpreted in the literature as a sign of a mismatch in the labor market; the fact that vacancies need to wait for new rounds of searching to be filled shows that workers in the unemployment stock did not have the abilities required by the open job positions.

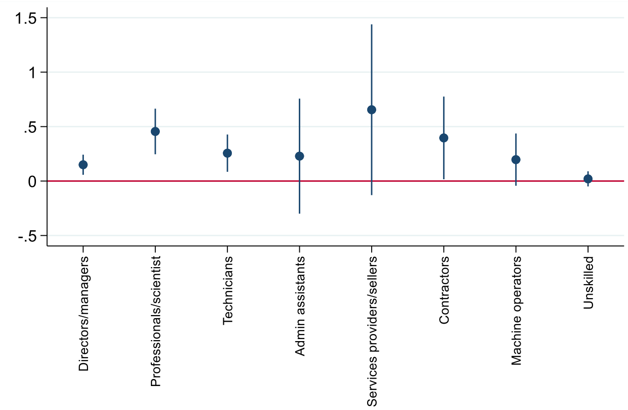
From the unemployed side, we find evidence of random matching for unskilled workers and professionals/scientists, as can be seen from the positive and significant coefficients α_s^D. In these occupations, the stock of unemployed is positive and significantly correlated with the formation of new matches; therefore, this significant correlation shows an informational lack of where the unemployed searching in these professions are located. The case of unskilled workers is somewhat particular because the only positive correlation with hires that we identify is with the stock of unemployment; this is the highest correlation among all occupations. Therefore, the stock of worker searchers in this profession is the most critical factor in determining new matches for unskilled workers. This finding is a sign of friction due to informational lack. The inflow of unskilled searchers could exacerbate congestion issues in this occupation; this is consistent with a negative correlation of this inflow with hires.

Graph 11: Estimated Coefficients for Formal Salaried Workers

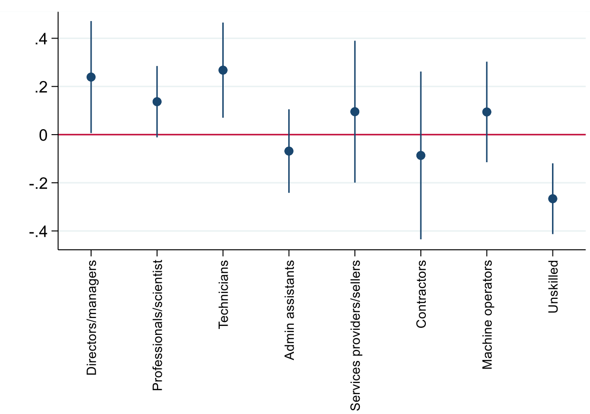
A. Coefficient γ_s^V: Flow of Vacancies



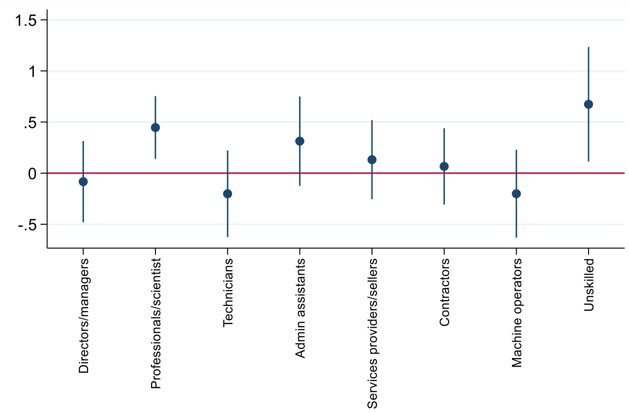
B. Coefficient α_s^V: Stock of Vacancies



C. Coefficient γ_s^D: Unemployment Flow



D. Coefficient α_s^D Unemployment Stock



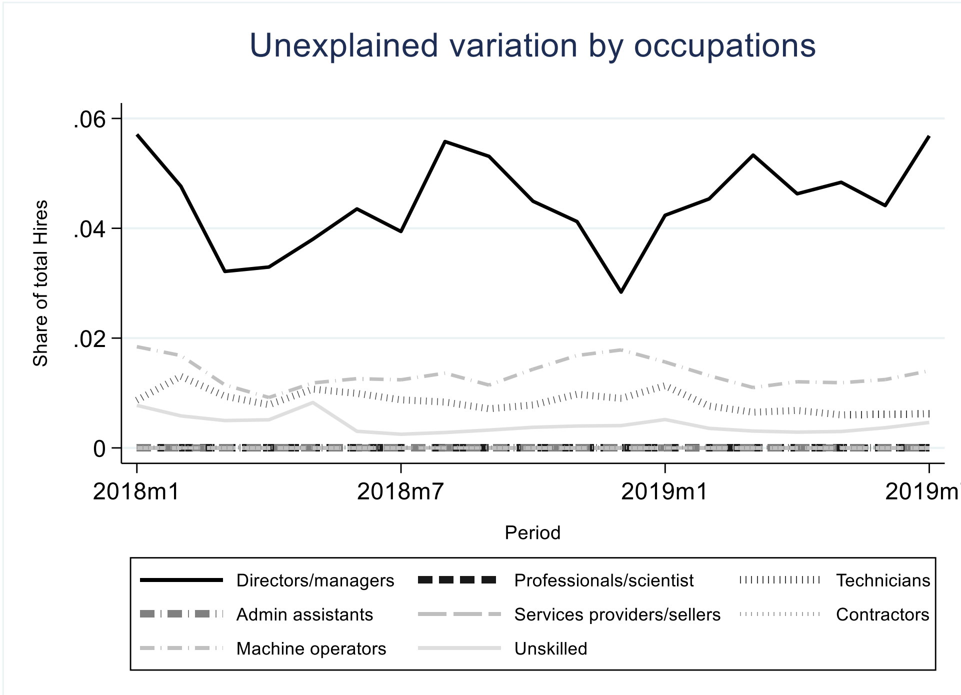
Notes: Graph 11 shows the plot of coefficients for $\gamma_s^V, \alpha_s^V, \gamma_s^D$ and α_s^D for each occupation, obtained from the estimation of equation (18). In this regression, we control by occupation, time-fixed effects, and linear and quadratic trend polynomials. Confidence intervals are constructed at 95%. Regression results are presented in Appendix E.

In light of the evidence presented in this sub-section, we can better understand the heterogeneity in labor market efficiency depicted by the BCs in sub-section 6.2 (see Graph 10).

From the side of job searchers, we find evidence of a mismatch in directors/managers, professionals/scientists, and especially technicians. In these cases, vacancies must wait for subsequent rounds to be filled. The evidence of the BC shows that labor markets for technicians are the most inefficient; its BC is the most distant to the origin. The inefficiency is explained by the mismatch between workers' abilities and the requirement of vacancies in this submarket. We identify informational lacks from the side of workers in the occupations of professionals/scientists, but especially for unskilled workers; in this latter case, we did not find evidence of a mismatch.

Finally, we comment on the residual component of the augmented matching function in equation (14); the parameter $\mu_{s,t}$ is a time-occupation varying intercept of this equation. It describes the residual variation of hires that is not explained by stocks/flows of vacancies or unemployment. We model this parameter as the summation of fixed effects: a fixed effect for each occupation, a fixed effect for each period, and a linear and quadratic trend for each occupation. In Graph 12, we present for each occupation the residual component of hires that is not explained by the augmented matching function as a proportion of total hires; this could be interpreted as a "rough" measure of efficiency in the matching process. For most occupations, this residual variation is not particularly important; nevertheless, some occupations might seem more efficient than others. Not surprisingly, directors/managers have the highest efficiency level from the point of view of this residual variation. From the BC evidence (see Graph 10), this occupation shows the highest efficiency level because it is the closest one to the origin. Other occupations as machine operators and technicians, show high levels of efficiency from the point of view of the residual variation. The latest one is also efficient from the perspective of the BC; its BC is one of the closest to the origin.

Graph 12: Unexplained variation of hires as a share of total hires



Notes: Graph 12 presents the residual component of hires not explained by the augmented matching function as a proportion of total hires for each occupation.

7. Using estimated vacancies to assess labor market tightness after the pandemic

The Covid-19 pandemic induced structural changes that significantly affected the Colombian labor market. In this context, and to assess the tightness of the Colombian labor market after the pandemic, we perform a similar exercise to the one presented in Domash and Summers (2022). The authors compare the actual unemployment rate with a predicted unemployment rate constructed from demand-side indicators, which they call the firm-side unemployment rate. In this application, we estimate a firm-side unemployment rate for the Colombian urban market. We regress the unemployment rate ($u_{a,t}$) in the metropolitan area a in period t as a function of the lags of the vacancy rate ($v_{a,t}$) and a proxy of the quits rate ($s_{a,t}$), as represented in the following equation.[3]

[3] The proxy for the quit rate is the replacement rate; it is the percentage of separations that are rehired within the three subsequent months after the separation took place. It is computed as indicated in Morales & Lobo (2020).

$$u_{a,t} = \alpha_{a,t} + \beta_{a,\tau} \sum_{\tau=0}^{L} \ln\left(v_{a,t-\tau}\right) + \delta_{a,\tau} \sum_{\tau=0}^{L} \ln\left(s_{a,t-\tau}\right) + \varepsilon_{a,t} \quad (19)$$

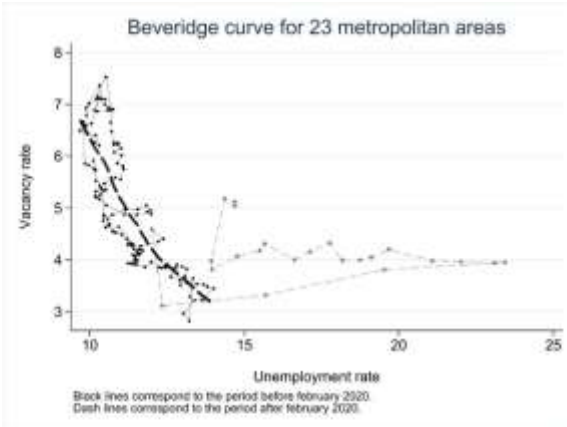
This regression is estimated until February 2020 to avoid changes in the structural relationship between unemployment and demand-side indicators that the pandemic might have caused. We used a linear-log model as Domash and Summers (2022) suggested and selected the polynomial lag length using the Bayesian information criterion (BIC). Finally, the period and MA fixed effects were included in the regression. The second step of this exercise uses the set of coefficients $\widehat{\beta_{a,\tau}}$, and $\widehat{\delta_{a,\tau}}$ in equation (19) to predict the firm-side unemployment rate in the post-pandemic period, given the vacancy and separation rates observed in this time.

We apply the methodology presented in section 3 to a panel with the level of hires for the 13 metropolitan areas from January 2008- June 2022, using the information on aggregated hires from the social security system's administrative records in Colombia.[4]. We decided not to use information from GEIH for this period because for some months during the pandemic, the statistical bureau in Colombia (DANE) did not use the complete questionnaire, and the questions that allow computing hires were not asked for several months.[5]. Since this subsection aims to assess the labor market tightness during the pandemic and the recovery period, we use information from the social security system, which is available for the whole period. Graph 13 presents an aggregate Beveridge curve for the 13 metropolitan areas using the estimated vacancies from GEIH. The dashed line corresponds to the period after the pandemic's beginning, February 2020. In this case, the Beveridge curve shows that the labor market is tightening, implying lower unemployment and higher vacancy rates.

[4] The administrative records correspond to the PILA (Spanish acronym: Planilla Integrada de Liquidación de Aportes). This data contains all payments to the payroll taxes made by firms and employees. From this information, a panel employer-employee might be constructed, from which traditional labor market flows as hires and separations might be constructed. Aggregated information from PILA is openly available from the ministry of health in Colombia.

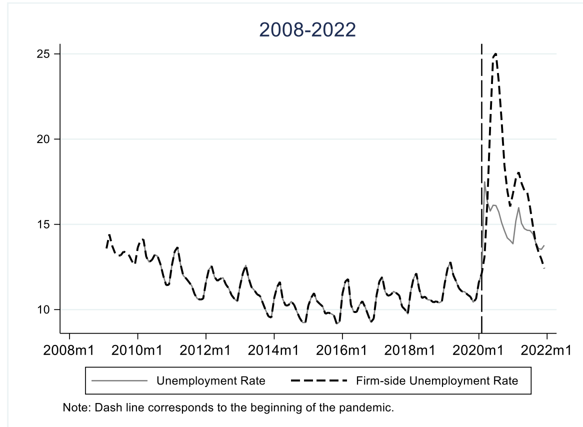
[5] The missing information can be interpolated or predicted.

Graph 13: Beveridge curve GEIH Estimated Vacancies



Notes: Graph 13 shows the Beveridge curve using PILA's estimated vacancies. The dashed line corresponds to the period after February 2020.

Graph 14: Actual Unemployment rate and Firm-side Unemployment rate



Notes: Graph 14 is based on the estimation of equation (19). The dark line represents the actual unemployment rate, and the gray line represents the firm-side unemployment rate.

Graph 14 presents the actual unemployment rate and the prediction of the firm-side unemployment rate based on the estimation of equation (19). The latter is higher than the former for the post-pandemic period, suggesting a tighter labor market from the demand side; for this period, the post-pandemic unemployment rate slack is higher using the actual unemployment rate than the firm-side unemployment rate. This finding suggests that the excess capacity illustrated by the standard unemployment rate is overestimated compared to what the demand-side perspective unemployment shows. In 2022, after a slow recovery process in the labor market, the standard unemployment rate and the firm side unemployment rate offer a signal of a similar level of slackness again.

8. Conclusion and Policy Implications

In this study, we developed a methodology that recovers an estimate of the average stock of vacancies using the information on aggregated hires for this purpose. The methodology constructs a mapping between vacancies and hires by exploiting the idea that those monthly vacancies are filled in the current and subsequent months. We use the information on total hires per economic sector in the US from JOLTS data to validate our estimations. Our predictions capture well the level and dynamics of aggregated vacancy stock. The observed level of vacancies is contained in the 95% confidence of the prediction for almost the entire study period. This methodology might be helpful in the case of developing countries with no quality data on vacancies; it can be easily implemented for any country since it uses input information from standard household surveys.

Using the methodology, we estimate vacancies for a set of occupations in Colombia. For each one of these submarkets, we describe Beveridge curves. From this evidence, we find that occupations with higher skills requirements as managers and professionals have a more efficient matching process than occupations as technicians, administrative assistants, machine operators, and other professions with tertiary education requirements, but not at the professional level. Furthermore, the formal markets for contractors and service providers are the ones that exhibited more significant levels of inefficiency in the matching process.

From the estimation of augmented matching functions, we can test the existence of mismatches or frictions due to informational lacks. From the side of job searchers, on the one hand, we find evidence of a mismatch in the occupations of directors/managers, professionals/scientists, and especially technicians. The evidence of the BC shows that the labor market for technicians is the most inefficient; the mismatch between the abilities of the workers and the requirement of the vacancies partially explains this inefficiency. Reducing frictions in this occupation will require education and job-oriented training policies. On the other hand, we identify informational lacks from the side of workers in the occupations of professionals/scientists, but especially for unskilled workers. The reductions of frictions in these cases will come from better intermediation and active search policies.

In a final application of our methodology, we use the predicted vacancies for calculating the firm-side unemployment rate; evidence suggests that the excess capacity illustrated by the standard unemployment rate is overestimated in comparison to what the demand-side perspective unemployment shows. This finding has been presented previously in the literature for the US labor market (Domash & Summers, 2022), and it has been explained as the results of temporary changes caused by Covid-19, structural changes in the age of the workforce, work incentives, and workers' reservation wages.

References

Abraham, K. G. (1983). Structural/Frictional vs. Deficient Demand Unemployment: Some New Evidence. *The American Economic Review*, *73*(4), 708–724. https://www.jstor.org/stable/1816568

Acemoglu, D., & Shimer, R. (1999). Holdups and efficiency with search frictions. *International Economic Review*, 40(4), 827-849.

Anderson, P. M., & Burgess, S. M. (2000). Empirical matching functions: estimation and interpretation using state-level data. *Review of Economics and Statistics*, 82(1), 93-102.

Andrews, M. J., Bradley, S., Stott, D., & Upward, R. (2013). Estimating the stock-flow matching model using micro data. *Journal of the European Economic Association*, 11(5), 1153-1177.

Arango, L. E. (2013). Puestos de trabajo vacantes según anuncios de la prensa escrita de las siete principales ciudades de Colombia. *Borradores de Economía*, *793*. http://www.banrep.gov.co/sites/default/files/publicaciones/archivos/be_793.xls

Barnichon, R. (2010). Building a composite Help-Wanted Index. *Economics Letters*, *109*(3), 175–178. https://doi.org/10.1016/j.econlet.2010.08.029

Beveridge, W. H. (1944). *Full Employment in a Free Society*, London (Allen).

Blanchard, O. J., Diamond, P., 1989. The Beveridge Curve. *Brookings Papers on Economic Activity*, 20(1), 1-60.

Coles, M. G. (1994). *Understanding the matching function: The role of newspapers and job agencies* (No. 939). CEPR Discussion Papers.

Coles, M. G., & Smith, E. (1998). Marketplaces and matching. *International Economic Review*, 239-254.

De Toledo, P. Á., Núñez, F., & Usabiaga, C. (2008). La función de emparejamiento en el mercado de trabajo español. *Revista de Economía Aplicada*, 16(48), 5-35.

Dmitrijeva, J. & Hazans, M. (2005). How do we match in Latvia? Does training Help?

Domash, A., & Summers, L. H. (2022). *How Tight are US Labor Markets?* (No. w29739). National Bureau of Economic Research.

Ebrahimy, E., & Shimer, R. (2010). Stock–flow matching. *Journal of Economic Theory*, 145(4), 1325-1353.

Elsby, M. W. L., Michaels, R., & Ratner, D. (2015). The Beveridge Curve: A Survey. *Journal of Economic Literature*, *53*(3). https://doi.org/10.1257/jel.53.3.571

Flórez, L. A., Morales, L. F., Medina, D., & Lobo, J. (2021). Labor flows across firm size, age, and economic sector in Colombia vs. the United States. *Small Business Economics*, 57(3), 1569-1600.

Lazear, E. P., & Spletzer, J. R. (2012). Hiring, churn, and the business cycle. *American Economic Review*, *102*(3), 575–579. https://doi.org/10.1257/aer.102.3.575

Moen, E. R. (1997). Competitive search equilibrium. *Journal of political Economy*, 105(2), 385-411.

Morales, L. F., & Lobo, J. (2020). Estimating vacancies from firms' hiring behavior: The case of a developing economy. *Journal of Economic and Social Measurement*, *45*(2), 139–170. https://doi.org/10.3233/JEM-210473

Morales, L. F., & Medina, D. (2019). Labour market fluidity and employment outcomes in Colombia: evidence from employer-employee linked data. *CEPAL Review.*

Morales, L. F., Ospino, C., & Amaral, N. (2021). Online Vacancies and its Role in Labor Market Performance. *Borradores de Economía*, *1174*. https://repositorio.banrep.gov.co/bitstream/handle/20.500.12134/10143/be_1174.pdf

Petrongolo, B., & Pissarides, C. A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic literature*, 39(2), 390-431.

Pissarides, C. A. (2000). *Equilibrium unemployment theory*. MIT press.

Sasaki, M. (2008). Matching Function For The Japanese Labour Market: Random Or Stock–Flow?. *Bulletin of Economic Research*, 60(2), 209-230.

Shimer, R. (2007). Mismatch. *American Economic Review*, 97(4), 1074-1101.US Bureau of Labor Statistics. (2020). *JOLTS handbook*.

US Bureau of Labor Statistics. (2020). *JOLTS handbook.* https://www.bls.gov/opub/hom/jlt/home.htm

Appendix A

In this appendix, we develop the intuitive relationship between the flow and the stock of vacancies. Since the vacancy stock includes the flow of new vacancies generated at period t and part of the flow of new vacancies generated in previous periods, the vacancy stock at the beginning of period t can be expressed as follows:

$$v_{j,s,t} = (1-\phi_0-\phi_1-\cdots-\phi_{R-1})\underline{v}_{j,s,t-R}+\cdots+(1-\phi_0-\phi_1)\underline{v}_{j,s,t-2} + (1-\phi_0)\underline{v}_{j,s,t-1}+\underline{v}_{j,s,t}$$

Similarly, the vacancy stock at the end of period t is given by:

$$v_{j,s,t} = (1-\phi_0-\phi_1-\phi_2-\cdots-\phi_R)\underline{v}_{j,s,t-R}+\cdots+(1-\phi_0-\phi_1-\phi_2)\underline{v}_{j,s,t-2} + (1-\phi_0-\phi_1)\underline{v}_{j,s,t-1}+(1-\phi_0)\underline{v}_{j,s,t}$$

We can create a system of equations by lagging R times the expression that represents the stock of vacancies at the end of period t:

$$v_{j,s,t} = (1-\phi_0)\underline{v}_{j,s,t}+(1-\phi_0-\phi_1)\underline{v}_{j,s,t-1}+(1-\phi_0-\phi_1-\phi_2)\underline{v}_{j,s,t-2}+\cdots + (1-\phi_0-\phi_1-\cdots-\phi_R)\underline{v}_{j,s,t-R}$$

$$v_{j,s,t-1} = (1-\phi_0)\underline{v}_{j,s,t-1}+(1-\phi_0-\phi_1)\underline{v}_{j,s,t-2}+(1-\phi_0-\phi_1-\phi_2)\underline{v}_{j,s,t-3}+\cdots + (1-\phi_0-\phi_1-\cdots-\phi_R)\underline{v}_{j,s,t-R-1}$$

$$v_{j,s,t-2} = (1-\phi_0)\underline{v}_{j,s,t-2}+(1-\phi_0-\phi_1)\underline{v}_{j,s,t-3}+(1-\phi_0-\phi_1-\phi_2)\underline{v}_{j,s,t-4}+\cdots + (1-\phi_0-\phi_1-\cdots-\phi_R)\underline{v}_{j,s,t-R-2}$$

$$\vdots$$

$$v_{j,s,t-R} = (1-\phi_0)\underline{v}_{j,s,t-R} + (1-\phi_0-\phi_1)\underline{v}_{j,s,t-R-1} + (1-\phi_0-\phi_1-\phi_2)\underline{v}_{j,s,t-R-2} + \cdots + (1-\phi_0-\phi_1-\cdots-\phi_R)\underline{v}_{j,s,t-2R}$$

These equations imply that part of the stock of vacancies in the current period corresponds to the flow of vacancies in the same period. If we solve the system for the current flow of each equation, we obtain:

$$\underline{v}_{j,s,t} = \frac{1}{(1-\phi_0)}\big[v_{j,s,t} - (1-\phi_0-\phi_1)\underline{v}_{j,s,t-1} - (1-\phi_0-\phi_1-\phi_2)\underline{v}_{j,s,t-2} - \cdots - (1-\phi_0-\phi_1-\phi_2-\cdots-\phi_R)\underline{v}_{j,s,t-R}\big] \approx \underline{\alpha}_0^s v_{j,s,t}$$

$$\underline{v}_{j,s,t-1} = \frac{1}{(1-\phi_0)}\big[v_{j,s,t-1} - (1-\phi_0-\phi_1)\underline{v}_{j,s,t-2} - (1-\phi_0-\phi_1-\phi_2)\underline{v}_{j,s,t-3} - \cdots - (1-\phi_0-\phi_1-\phi_2-\cdots-\phi_R)\underline{v}_{j,s,t-R-1}\big] \approx \underline{\alpha}_1^s v_{j,s,t-1}$$

$$\underline{v}_{j,s,t-2} = \frac{1}{(1-\phi_0)}\big[v_{j,s,t-2} - (1-\phi_0-\phi_1)\underline{v}_{j,s,t-3} - (1-\phi_0-\phi_1-\phi_2)\underline{v}_{j,s,t-4} - \cdots - (1-\phi_0-\phi_1-\phi_2-\cdots-\phi_R)\underline{v}_{j,s,t-R-2}\big] \approx \underline{\alpha}_2^s v_{j,s,t-2}$$

$$\vdots$$

$$\underline{v}_{j,s,t-R} = \frac{1}{(1-\phi_0)}\big[v_{j,s,t-R} - (1-\phi_0-\phi_1)\underline{v}_{j,s,t-R-1} - (1-\phi_0-\phi_1-\phi_2)\underline{v}_{j,s,t-R-2} - \cdots - (1-\phi_0-\phi_1-\phi_2-\cdots-\phi_R)\underline{v}_{j,s,t-2R}\big] \approx \underline{\alpha}_R^s v_{j,s,t-R}$$

Note that, in the right-hand side of each equation, the flows only correspond to periods behind the period in question. Replacing each flow term in equation (1), we can find an expression for hirings in period t as a function of the stock of vacancies:

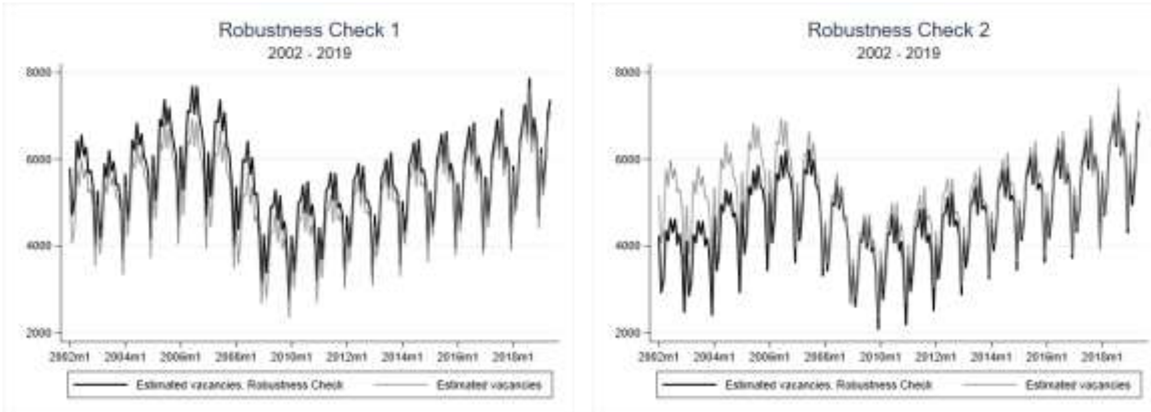
$$h_{j,s,t} = \phi_0^s \underline{\alpha}_0^s v_{j,s,t} + \alpha_1^s \underline{\alpha}_1^s v_{j,s,t-1} + \cdots + \underline{\alpha}_R^s v_{j,s,t-R}$$

And finally, this expression can be easily transformed into equation (2):

$$h_{j,s,t} = \alpha_0^s v_{j,s,t} + \alpha_1^s v_{j,s,t-1} + \cdots + \alpha_R^s v_{j,s,t-R}$$

Appendix B: Robustness Checks

Quadratic trend, 4 year window **Linear trend, 6 year window**



3Notes: In this figure, we present the robustness checks that compare the aggregated vacancies from the original specification and two new predicted series of the stock of vacancies, computed using the methodology presented in section 3. Panel A estimates the regression controlling by occupation and month fixed effects; and by the interaction between occupation and year, and linear and quadratic trends for occupations, using 4 years of window. Panel B estimates the regression controlling by occupation and month fixed effects; and by the interaction between occupation and year, and linear trends for occupations, using 6 years of window.

Appendix C

	Hires Window: 2016-2019 Best lag: 4			
	L1.Hires	**L2.Hires**	**L3.Hires**	**L4.Hires**
Sector				
1. Mining and logging	-0.524	-2.094	-0.0342	3.648
	(5.132)	(5.411)	(5.625)	(5.131)
2. Construction	0.307**	-0.0945	0.144	-0.0217
	(0.134)	(0.149)	(0.185)	(0.124)
3. Durable goods manufacturing	-0.134	-0.632*	0.366	0.917***
	(0.321)	(0.346)	(0.352)	(0.349)
4. Nondurable goods manufacturing	-0.213	-1.241*	0.188	1.105*
	(0.586)	(0.644)	(0.669)	(0.588)
5. Wholesale trade	-0.219	-1.157**	0.124	0.428
	(0.477)	(0.505)	(0.527)	(0.534)
6. Retail trade	0.278***	-0.404***	-0.0698	-0.180***
	(0.0655)	(0.0782)	(0.0731)	(0.0695)
7. Transportation, warehousing and utilities	-0.296	-0.138	-0.495*	-0.493**

	Hires			
	Window: 2016-2019			
	Best lag: 4			
	L1.Hires	**L2.Hires**	**L3.Hires**	**L4.Hires**
	(0.296)	(0.273)	(0.274)	(0.228)
8. Information	-1.020	-1.863	-0.265	0.758
	(1.266)	(1.239)	(1.344)	(1.448)
9. Finance and insurance	-0.499	-0.923**	0.352	0.612
	(0.404)	(0.439)	(0.469)	(0.477)
10. Real estate and rental and leasing	-0.508	-0.381	1.683	2.428**
	(1.098)	(1.341)	(1.540)	(1.207)
11. Professional and business services	0.159***	-0.405***	0.125*	-0.308***
	(0.0541)	(0.0602)	(0.0676)	(0.0591)
12. Educational services	0.112	-0.435	-0.833	-0.191
	(0.514)	(0.674)	(0.665)	(0.514)
13. Health care and social assistance	0.132	-0.403***	-0.0736	-0.520***
	(0.0846)	(0.0834)	(0.0898)	(0.0896)
14. Arts, entertainment, and recreation	0.763**	-0.194	0.0420	0.506
	(0.388)	(0.502)	(0.584)	(0.385)
15. Accommodation and food services	0.380***	0.162	0.108	-0.310***
	(0.0978)	(0.125)	(0.125)	(0.0912)
16. Other services	0.161	-0.310	0.228	0.396
	(0.231)	(0.257)	(0.288)	(0.242)
17. Federal	-2.183	-3.223	-0.854	0.0117
	(2.073)	(2.049)	(2.097)	(2.091)
18. State and local	0.304***	-0.156**	-0.136**	-0.371***
	(0.0642)	(0.0685)	(0.0676)	(0.0687)
Observations	774			
R-squared	0.984			

Notes: * significant at 10%; ** significant at 5.0%; *** significant at 1.0%. This table shows the estimation of the equation (10) for the optimal polynomial length in the window 2016-2019. In this case, the selected model included 4 lags of the variable hire.

Appendix D

Occupation	Obs.	Mean	Median	Std.	Min	Max
Employment						
1. Directors/managers	156	635,863	638,626	65,137	454,583	821,848
2. Professionals/scientist	156	1,252,788	1,243,560	170,473	928,904	1,610,938
3. Technicians	156	542,309	566,219	112,082	331,483	742,881
4. Admin assistants	156	1,022,291	1,032,319	130,863	618,290	1,263,973
5. Services providers/sellers	156	4,144,820	4,296,764	417,327	3,231,011	4,697,535
6. Contractors	156	1,978,695	1,995,183	133,563	1,663,079	2,273,552
7. Machine operators	156	1,034,703	1,064,916	120,857	743,237	1,344,875
8. Unskilled	156	286,454	284,556	25,542	211,469	362,253
Formal salaried workers						
1. Directors/managers	156	351,855	347,127	53,327	245,512	518,358
2. Professionals/scientist	156	745,885	721,464	117,893	542,814	1,004,251
3. Technicians	156	346,902	354,359	79,630	198,652	505,875
4. Admin assistants	156	774,984	772,388	124,307	506,873	1,018,248
5. Services providers/sellers	156	1,159,862	1,160,522	183,759	841,195	1,487,507
6. Contractors	156	563,653	557,691	62,059	408,079	713,357
7. Machine operators	156	406,021	416,408	66,395	276,960	565,718
8. Unskilled	156	111,075	109,515	15,629	72,680	154,470
Unemployed						
1. Directors/managers	156	38,991	37,970	8,914	19,478	63,543
2. Professionals/scientist	156	157,660	150,439	34,262	99,072	285,929
3. Technicians	156	60,980	59,840	15,872	27,485	110,054
4. Admin assistants	156	199,556	198,650	38,849	95,507	314,593
5. Services providers/sellers	156	573,159	568,999	56,377	462,817	716,784
6. Contractors	156	210,626	207,723	38,189	114,813	310,367
7. Machine operators	156	100,759	100,743	17,917	57,540	163,682
8. Unskilled	156	30,953	28,252	10,828	11,624	64,440
Short unemployment						
1. Directors/managers	156	9,828	8,954	4,255	1,326	24,171
2. Professionals/scientist	156	41,331	36,672	19,745	14,435	118,282
3. Technicians	156	18,499	17,397	7,978	5,041	50,043
4. Admin assistants	156	59,947	55,774	20,609	26,417	144,802
5. Services providers/sellers	156	191,114	186,143	40,404	117,279	373,823
6. Contractors	156	88,613	83,379	21,352	45,149	171,786
7. Machine operators	156	34,561	32,308	9,983	17,943	72,366
8. Unskilled	156	10,071	9,142	4,446	2,860	28,044

Occupation	Obs.	Mean	Median	Std.	Min	Max
Hires						
1. Directors/managers	156	8,690	8,195	4,219	934	23,875
2. Professionals/scientist	156	26,842	23,387	13,987	6,284	87,614
3. Technicians	156	18,674	17,584	7,950	3,567	38,629
4. Admin assistants	156	41,538	41,490	12,337	17,035	75,005
5. Services providers/sellers	156	67,101	66,131	19,457	31,663	126,765
6. Contractors	156	50,640	49,795	14,688	19,558	91,464
7. Machine operators	156	24,947	24,979	8,895	6,560	48,065
8. Unskilled	156	6,670	6,150	3,727	425	20,578

Notes: This table summarizes the descriptive statistics by occupation of the variables used in the methodology application. The sample corresponds to a panel for 8 occupations with monthly data between January 2007 and December 2019.

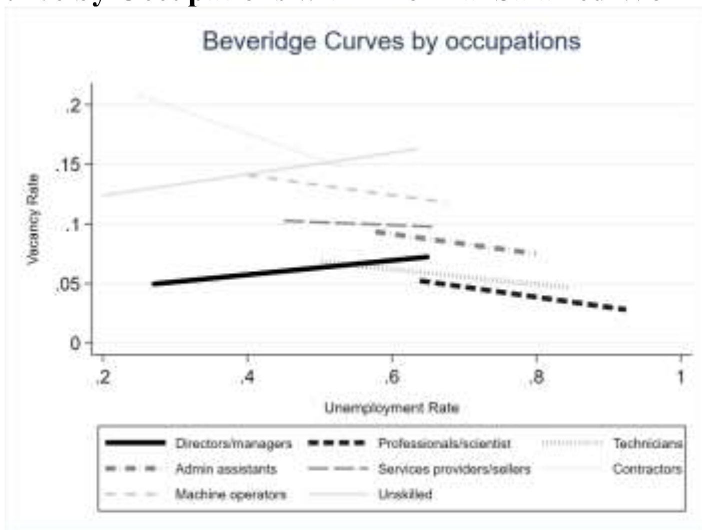
Appendix E

	Hires				
Formal Salaried Workers	Stock of Vacancies γ_s^V	Lag of Stock of Vacancies $(\alpha_s^V - \gamma_s^V)$	Lag of Hires $\beta_0^s(\gamma_s^V - \alpha_s^V)$	Stock of Unemployed α_s^D	Flow of unemployed γ_s^D
1. Directors/managers	0.1499***	0.0415	0.0195	-0.0833	0.2390**
	(0.0453)	(0.0513)	(0.0357)	(0.1948)	(0.1140)
2. Professionals/scientist	0.4554***	0.1958	0.1195	0.4463***	0.1368*
	(0.1026)	(0.1725)	(0.1434)	(0.1505)	(0.0727)
3. Technicians	0.2561***	0.2650**	-0.0664	-0.2016	0.2677***
	(0.0839)	(0.1091)	(0.0994)	(0.2079)	(0.0969)
4. Admin assistants	0.2288	0.5628	-0.0758	0.3132	-0.0682
	(0.2589)	(0.3570)	(0.1056)	(0.2143)	(0.0850)
5. Services providers/sellers	0.6549*	0.2018	0.0314	0.1321	0.0953
	(0.3845)	(0.3551)	(0.0657)	(0.1898)	(0.1444)
6. Contractors	0.3961**	0.7072***	-0.1919	0.0664	-0.0863
	(0.1862)	(0.2340)	(0.1314)	(0.1833)	(0.1708)
7. Machine operators	0.1968	0.4482***	-0.0939	-0.2011	0.0942
	(0.1176)	(0.1535)	(0.0595)	(0.2108)	(0.1024)
8. Unskilled	0.0200	0.0224	0.0823***	0.6738**	-0.2661***
	(0.0340)	(0.0493)	(0.0246)	(0.2747)	(0.0720)
Observations	1,104				
R-squared	0.8984				
Occupation-FE	Yes				
Time-FE	Yes				
Trend*Occupation	Yes				

Notes: * significant at 10%; ** significant at 5.0%; *** significant at 1.0%. This table shows the estimation of the equation (18). In this regression, we control by occupation and time fixed effects and by linear and quadratic trend polynomials.

Appendix F.

Beveridge curve by Occupations with Informal Salaried Worker Vacancies



Notes: This figure shows the linear estimation of the Beveridge curve for each occupation. We find that managers and professionals are the most efficient informal labor markets. In contrast, unskilled workers and contractors are the most inefficient ones.


BANCO DE LA REPÚBLICA
COLOMBIA