

ORIGINAL ARTICLE

Do job vacancies variations anticipate employment variations by sector? Some preliminary evidence from Italy

Pietro Giorgio Lovaglio

Department of Statistics and Quantitative Methods, University Bicocca-Milan (Italy), Milan, Italy

Correspondence

Pietro Giorgio Lovaglio, Department of Statistics and Quantitative Methods, University Bicocca-Milan (Italy), Via Bicocca degli Arcimboldi 8, 20126 Milan, Italy.

Email: piergiorgio.lovaglio@unimib.it

Abstract

Government policy has placed increasing emphasis on the need for robust labour market projections. The job vacancy rate is a key indicator of the state of the economy underpinning most monetary policy decisions. However, its variation over time is rarely studied in relation to employment variations, especially at the sectoral level. The present paper assesses whether changes in the number of vacancies from quarter to quarter are a leading anticipator of employment variation in certain economic sectors over the previous decade in Italy, using multivariate time-series tools (the vector autoregressive and error correction models) with Eurostat data. As robustness checks for integration order and cointegration, we compare traditional critical values with those provided by response surface models. To the best of our knowledge, no previous study has evaluated this relationship using Italian data over the last decade. The results demonstrate that percentage changes in numbers employed (occupied persons) react to percentage changes in vacancies (one-quarter lagged), but not vice versa, indicating that variations of vacancies are weakly exogenous. The fastest short-term adjustment from disequilibrium is seen in the construction industry, whereas the manufacturing and the information

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *LABOUR* published by Fondazione Giacomo Brodolini and John Wiley & Sons Ltd.

and communication technology sectors demonstrate the strongest long-run relationships among variations. This suggests that the matching rates – the likelihood that a vacancy is filled – are higher for these than for other sectors, as a result of developments in recruitment technology for professional figures of such industries.

KEYWORDS

cointegration, job vacancy rate, labour demand, long-run equilibrium

JEL CLASSIFICATION

J21; J23; C22

1 | INTRODUCTION

Government policy has placed increasing emphasis on the need for robust labour market stock projections to be made freely available at a detailed occupational and sectoral level to assist in policy and planning for the provision of education and training. From this perspective, recent EU labour market policies indicate that job vacancy data can be exploited to improve the knowledge and functioning of the labour market by matching supply and demand more closely (Cedefop, 2017; OECD, 2018).

Data on vacancies are usually mentioned in the context of either labour market tightness (where the demand for labour exceeds the supply) or that of labour market slack. As the number of vacancies changes in line with the number of jobs, a helpful way of looking at the data is in terms of the job vacancy rate (JVR). This is the ratio of total job vacancies to the sum of occupied jobs and job vacancies (Cedefop, 2017; Eurostat, 2018). A higher JVR reflects a demand for labour that is not satisfied, whereas a lower rate can indicate a skills mismatch related to situations in which employers are unable to fill open jobs in specific periods or economic sectors. The JVR is one of the key indicators monitored by the European Commission and the European Central Bank in their assessment of the business cycle and of the degree of tightness in the labour market (Eurostat, 2019).

In the economic literature, the JVR is often used to assess the educational mismatch between vacancies and unemployment (Dur, 1999; Gorter & van Ours, 1994; Jackman & Roper, 1987). The ratio is also used to assess its relation to the unemployment rate in the context of the Beveridge curve. This approach has been used to assess the efficiency of the labour market, in particular the mismatch between unemployment and vacancies, in countries such as the Netherlands (Gorter & van Ours 1994), the United States (Blanchard & Diamond, 1989) and the United Kingdom (Jackman et al., 1989), as well as in different sectors (see Elsby et al., 2015 for a review). More recently, researchers have used JVR data from web sources to predict the rate of unemployment (D'Amuri & Marcucci, 2009; Askitas & Zimmermann, 2009; Fondeur & Karamé, 2011). Other studies combine online and official supply-side data – mainly, the European Union's Labour Force Survey (LFS) – to compare the sectoral and occupational distribution of online job vacancies (Štefánik, 2012; Steinmetz et al., 2009) to better establish the correct coverage of web sources.

On the supply side, institutions publish forecasts of the size and composition of the population and trends in labour force participation rates and employment, based on extrapolations from LFS data. Prediction methods range from well-known time-series models (see, US Bureau of Labor Statistics, 2019; ILO, 2017; Australian Government, 2019) to complex econometric modeling that exploits relationships between output, labour demand and labour supply (Wilson et al., 2016, for the UK).

Despite the importance of job vacancies as a key indicator of the state of the economy and one that underpins most monetary policy decisions, variations in job vacancy counts (or rates) are rarely studied or used to forecast variations in employment. Furthermore, details at the sectoral level provide a limited assessment of the actual underlying labour market conditions. Nevertheless, there are several reasons that a deep assessment of developments in vacancies is useful for economic analysis, not least of which is that vacancies may offer leading indicator properties for employment developments.

The mechanism explaining the increase in employment in a given period is complex, depending on the number of job vacancies that are filled in the period (affected by the matching efficiency of the labour market and changes in recruitment technology) and on the number of people that leave a job for any reason or turnover. Nevertheless, indicators of labour demand are commonly used as inputs into near-term forecasts of employment growth.

In the short term, the demand for labour is affected by growth in (firm) output – this may lead to an initial increase in the number of hours worked by existing employees and later to an increase in the number of workers employed. A job vacancy then exists between the period of increased output and the eventual employment of additional staff. As there may be some time between the start of the recruitment activity and the position being filled, an increase in job vacancies in the current period may lead to an increase in employment in a subsequent period. Similarly, there may be a fall in demand for labour in the form of a slowing of search activity before the employment evolution becomes apparent.

Previous studies examine job vacancy data as a possible leading indicator of cyclical employment dynamics (Haggard-Guénette, 1989; Zagorsky, 1998; Amoah, 2000; European Central Bank, 2002; Australian Bureau of Statistics, 2003; Valletta, 2005; Ruth et al., 2006; Mandrone et al., 2010). Econometric estimates suggest that the addition of lagged values of vacancy stocks to simple autoregressive models or multivariate time-series models improves their explanatory power in relation to employment growth.

Specifically, the literature on vacancies as a leading indicator finds that employment growth is more highly correlated with changes in vacancy stocks over time than with the actual level of vacancies at any point in time.¹ This suggests that changes in the number of vacancies from quarter to quarter (or quarterly seasonal changes) may be an early sign of changes in labour demand while potentially remaining relatively unaffected by structural changes and developments in recruitment technology.

Results indicate that quarterly growth in job vacancies, rather than exhibiting contemporaneous movement, leads employment growth by between one- and three-quarters. These results suggest that deeper analysis of the co-movement of two series could contribute to the understanding of trends in filled and unfilled job demand in the labour market, as well as identifying sectors or areas at risk of human resource shortages.

In the present paper, we evaluate the empirical relationships over time between variation in quarterly stocks of occupied posts and vacant posts in Italy from Q1 2011 to Q3 2018 by specific sectors of economic activity. We use official statistical sources for the number of job vacancies

and the number of employed people (the Eurostat Vacancies Survey and the LFS, respectively). Specifically, the aim of this study was to assess the following issues:

1. whether variations in job vacancies move longitudinally in line with (actual and lagged) variations in numbers employed and
2. whether variations in job vacancies are significant anticipators (in the sense of Granger causality) of variations in numbers employed and, therefore, useful predictors in forecasting future employment outcomes in specific economic sectors.

The strength of the relationship between the two series will, in practice, allow us to evaluate empirically the extent to which vacancy growth can be considered a leading indicator of employment growth in future (or previous) quarters – a lead-lag relationship. To the best of our knowledge, no previous study estimates this relationship using Italian aggregate and sectoral data over an extended period.

Methodologically, we use the classical time-series tools of vector autoregression (VAR: Sims, 1980) and vector error correction (VECM: Enders, 2010) models also assessing other issues (stationarity, trend, seasonality and cointegration) to evaluate the correct strategy, VECM (long-run relation with cointegrated rather than stationary series) or VAR (short-run relation, with stationary or difference stationary variables). Robust critical values based on surface regression and cross-correlation analyses among lagged series are performed as robustness checks. The analysis is repeated by activity sector to cover those sectors most representative of the Italian economy in terms of both occupied and vacant posts.

The remainder of the paper is structured as follows. Section 2 explains the data sources and the series analysed. In Section 3, we detail our strategy and the time-series concepts used. Section 4 presents the results, and in Section 5, we draw our conclusions.

2 | JOB VACANCY AND EMPLOYMENT DATA SOURCES

For the purposes of official national statistical data, a job vacancy is defined as a paid post that is newly created, unoccupied or about to become vacant, (a) for which the employer is taking active steps and is prepared to take further steps to find a suitable candidate from outside the enterprise concerned, and (b) which the employer intends to fill either immediately or within a specific period (Eurostat, 2018). Steps taken to fill paid posts include providing notice of the vacancy to the public employment services, private employment agencies and/or headhunters, and advertising in the media such as on the Internet and in newspapers and magazines.

The Job Vacancy Survey (JVS), published by Eurostat and compulsory for all Member States from January 2010, represents the stock of vacant posts at a given reference date in a quarter.

European regulations state that the quarterly JVS covers the population of firms with at least one employee in all areas of economic activity, with the exception of households and extraterritorial organizations or bodies. The survey is optional for agriculture firms. The Italian survey covers only firms with at least 10 employees. Data for job vacancies are collected separately for enterprises with 10 to 499 employees and more than 500 employees. For enterprises with 10 to 499 employees, data are collected by the National Statistics Institute as part of their quarterly survey on job vacancies and hours worked (VELA). For enterprises with more than 500 employees, data are collected through an extended quarterly version of a well-established monthly census survey of large enterprises concerning employment, hours worked, wages and labour costs (Istat, 2018).

Data on job vacancies represent the stock of vacant posts on a given reference date (the last day of a given quarter). The quarterly data are broken down by economic activity (at the section level) in accordance with the revised statistical classification of economic activities in the European Community (NACE Rev. 2).² Data on those employed are the official Labour Force Survey (LFS) data for Italy, provided by Eurostat.

The LFS is a large household sample survey that provides quarterly data on the labour participation of people aged 15 and over. It is a continuous survey conducted by national statistical institutes across Europe and processed centrally by Eurostat, providing information on the main labour market indicators (including employment status, characteristics of primary employment, work experience and job search activity) and broken down according to the main socio-demographic variables. The employment level is defined as the number of people engaged in productive activities in the economy. An employed person is a person aged 15 and over who during the reference week performed work for pay, profit or family gain, even if only for one hour.

We used Eurostat's public quarterly cross-sectional data for Italy, where the numbers of people employed are disaggregated by economic activity (see the table 'Employment by sex, age and economic activity', https://ec.europa.eu/eurostat/databrowser/view/lfsq_egana/default/table?lang=en). We merged LFS and vacancy data by quarter and sector (according to the NACE codes) to identify the quarters with vacancies, the total number of vacancies and employed persons at the end of the quarter for each sector. Because job vacancies are not reported for three sectors (agriculture, forestry and fishing; real estate activities; and public administration and defence/compulsory social security) for Italy, we omitted these from both series for the analysis.

3 | METHODOLOGICAL APPROACH

The time series of those employed (count data) is persistent in nature, and its order of magnitude (millions) is clearly incomparable with that of vacancies (thousands). Moreover, both time series are characterized by a complex and unknown data-generating process, with a strong seasonal/quarterly pattern and possible non-stationarity. This complexity should be carefully evaluated, particularly when disaggregating each series by sector.

To this end, we analyse the so-called quarterly seasonal natural log returns (henceforth, seasonal returns), leading the series of percentage changes in numbers employed and vacancies with respect to the same quarter in the preceding year.

In the series of counts for numbers employed (Occ_t) and vacancies (Jwc_t), in quarter t (Q1 2011 to Q3 2018), seasonal returns are written as $o_t = \log(Occ_t/Occ_{t-4}) \times 100$ and $w_t = \log(Jwc_t/Jwc_{t-4}) \times 100$. Seasonal differencing is a crude form of seasonal adjustment that – at the cost of one year of data (the four quarters of 2011) – removes the seasonal component and can also eliminate the seasonal non-stationarity of the random walk type. However, if a stochastic or deterministic trend is present in the data, it is necessary to assess whether there are other issues of non-stationarity (due to unit roots or a deterministic trend) such as by establishing the order of integration, $I(d)$, of the series.

We test for the presence of a unit root using the Augmented Dickey–Fuller (ADF) test (Said & Dickey, 1984), specifying an appropriate lag length based on the statistical significance of the lagged coefficients, following a general-to-specific strategy (Enders 2010). This strategy is based on phi statistics resting on joint multivariate tests for unit roots and the presence of a deterministic trend (ϕ_3), a unit root with drift (ϕ_2) or a unit root only (ϕ_1). Possible stationarity and order

of integration $I(d)$ are also assessed using the auto.arima algorithm (Hyndman & Khandakar, 2008). This algorithm returns the best autoregressive integrated moving average (ARIMA) model for univariate time series by minimising the corrected Akaike information criterion (AICc). The order of integration is found using Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests (Hyndman & Athanasopoulos, 2017).

It is difficult to establish the exact role of endogenous or exogenous factors for the count of employed persons and vacancies. Therefore, for the relationships among series, our approach is to treat both series symmetrically using both VAR and VECM models to analyse series co-movements (in terms of direction, speed and statistical significance). A detailed description of both models can be found in Appendix 1.

To summarize our empirical strategy, where both o_t and w_t are stationary, we estimate their relationship in levels using a VAR model. Where both series are $I(1)$ and cointegrated, we estimate a VECM model, using the Engle & Granger (1987) (EG) two-step strategy, normalising the first step equation on o_t .

If the series are not cointegrated, we difference them (Δo_t and Δw_t) and estimate a VAR in first differences. These differenced series can be interpreted as the amount by which the change from the previous to the current period ($x_t - x_{t-1}$) is different from the change over the same period observed exactly one year earlier ($x_{t-4} - x_{t-5}$); see Appendix 1 for details. Because our series are relatively short, for parsimony in each equation of the VAR and VECM models, we specify only one lag for both variables o_t and w_t (or Δo_t and Δw_t).

The main coefficients of interest in the estimated models are the error correction term for the VECM model (the long-run relationship between series in levels) and the lagged coefficients of exogenous variables in the VAR model and the VECM (short-run relationships in differences), which provide an estimate of the direction and significance of Granger causality, between the two series or their differenced version.

Where there was a different stationarity order d for o_t and w_t , no further analyses were undertaken (Enders, 2010; Stewart, 2011). We begin using aggregated series (by economic sector) and then replicate our analysis for each sector.

3.1 | Robustness of integration order and Engle–Granger strategy

Despite its popularity, the EG 2-step methodology may be inefficient. OLS estimator in the static equation, although consistent, may be substantially biased for small samples (T), partly due to serial correlation in the residuals. Moreover, the critical values do not follow conventional distributions and testing for cointegration requires non-standard critical values; these are usually calculated using Monte Carlo simulations. Engle and Granger (1987) tabulate a limited set of critical values for Dickey–Fuller and Engle–Granger tests (see their Table 2), which is further extended by MacKinnon (1991, 1996).

The same holds for the likelihood ratio test of Johansen (1988, 1991), based on the rank of a cointegrated matrix using parametric estimation of a VAR model, for which several authors simulate asymptotic distributions of test statistics. Several studies suggest, however, that asymptotic critical values may be misleading in small samples. Johansen (1988), Johansen and Juselius (1990) and Osterwald-Lenum (1992) include critical values for the Johansen procedure under typical assumptions concerning deterministic terms and the number of stochastic variables (N). Gregory (1994), Reinsel and Ahn (1988) and Cheung and Lai (1993) show that for the Johansen tests, the rejection frequency when the null hypothesis of no cointegration is true is significantly

higher for small samples, and the critical values should be adjusted upwards (or the test statistics downwards). To this end, Johansen (1995), Doornik (1998) and MacKinnon et al. (1999) provide more accurate estimates of these critical values.

In addition to the EG and Johansen procedures, the single-equation conditional error correction model offers another approach to testing whether non-stationary economic time series are cointegrated. Banerjee et al. (1993, 1998) tabulate critical values for an error correction model with two variables at different sample sizes. Harbo et al. (1998), MacKinnon et al. (1999), Pesaran et al. (2000) and Ericsson and MacKinnon (2002) provide asymptotic critical values for single- and multiple-equation error correction models. Engle and Yoo (1987), Phillips and Ouliaris (1990) and MacKinnon (1996) provide critical values for different versions of the EG test for various values of variables, sample sizes and deterministic terms. The critical values reported in the literature often only cover a small subset of the possible model specifications and sample sizes. Nevertheless, they are prone to experimental error due to a smaller number of replications in the respective Monte Carlo simulations.

MacKinnon (2010) estimates both finite-sample and asymptotic distribution functions for the cointegration t -statistic of the EG procedure (that comprise the Dickey–Fuller statistic as a special case) by estimating response surface models that predict the quantiles of the distributions – critical values (CVs) – as a function of the sample size, and the number of deterministic terms and variables in the cointegration.

As a robustness check, we compare traditional tabulated critical values for the order of integration (ADF phi statistics) and assessment of cointegration (ADF test on the residuals of the static equation from the EG procedure) with those provided by MacKinnon (2010), using finite-sample and asymptotic critical values. As a robustness check for stationarity and the order of integration of the univariate series, as mentioned, we also consider the best ARIMA model selected using the AICc statistics (Hurvich & Tsai 1989), implemented in the `auto.arima` procedure; a corrected version of the AIC is also useful to select an adequately parsimonious model for small samples.

4 | RESULTS

In the period Q1 2011 to Q3 2018, Eurostat collected a total of 1,856,455 job vacancies for Italy across various sectors – excluding the agriculture, real estate and public administration sectors. The vacancies are distributed almost equally across quarters (with an average of 60,000 vacancies per quarter over 31 quarters), although from Q1 2017, they show a steady overall increase (Figure 1).

Regarding the distribution of vacancies by sector, we find manufacturing, retail, accommodation/food and professional, scientific and technical activities account for 63 percent of all vacancies. (Table 1 reports the distribution for the last available quarter of data.)

Notably, the percentage of vacancies reflects the structure of employment, with the exception of some overrepresentation, as expected, of vacancies in information and communications technology (ICT), manufacturing, and professional and administrative support-service activities. Differences in the sector's shares of vacancies and employment may be the result of skill shortages (unmet demand for high skill occupations typically emerges in vacancy data, see Garasto et al., 2021). However, these differences may also reflect variations in turnover rates between sectors. Sectors with vacancies that are over-represented in employment numbers might generally have higher turnover rates.

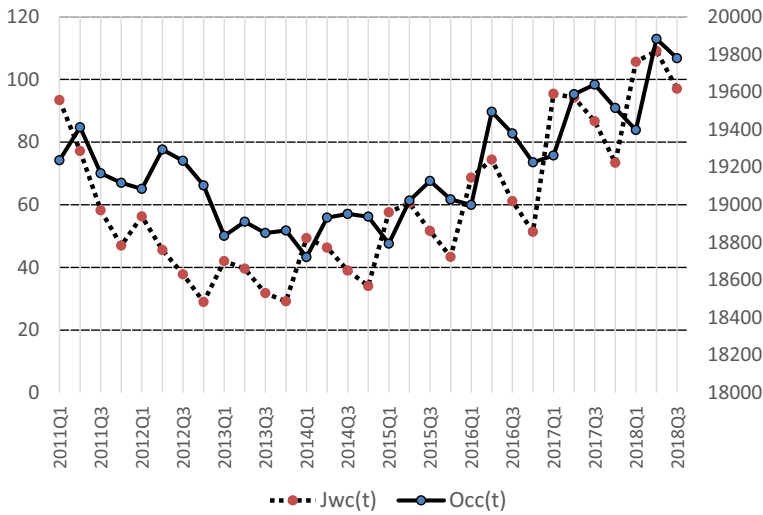


FIGURE 1 Quarterly LFS occupied counts (Occ) and job vacancies counts (Jwc), 2011Q1–2018Q3 (thousands).

4.1 | Modelling aggregate sectors

Using aggregated series over the various sectors, Figure 2 shows the analysed seasonal returns (o_t and w_t). We remove strong seasonality and find both series appear to be non-stationary: For the o_t ADF test, where no significant lags are found, the tau statistic is -1.611 (in a model with one unit root without either trend or drift, according to the phi statistics), which compared with the critical value of -1.950 (at the 5 percent level of significance), warrants the conclusion that the series has a unit root (stochastic trend). We draw the same conclusion for the series w_t ($\tau = -1.256$).

First, differencing of both series (Δo_t , Δw_t) results in stationary series. The finite-sample ($T = 26$) MacKinnon (2010, Table 1 for $N = 1$) and critical value (-1.954 at the 5 percent level of significance and -1.939 , the asymptotic critical value) confirm that both series have one unit root and are $I(1)$.

The integration order of the o_t and w_t series is confirmed by auto.arima. We find the best model is $(2,1,0)(1,0,0)_4$ for o_t and $(0,1,0)(0,0,1)_4$ for w_t , indicating unit roots in the non-seasonal part of the process but stationarity in the seasonal part (indexed as subscript 4). Both models are reasonable representations, based on the inspection of autocorrelation functions (ACF) and partial autocorrelation functions (PACF), and this is confirmed by all residual diagnostics (no significant spikes for ACF and PACF; all p -values of Ljung–Box statistics > 0.05 ; the qq-plot of the standardized residual endorses normality). In particular, Δw_t is modelled as a pure seasonal moving average, whereas Δo_t is modelled as a combination of (seasonal and non-seasonal) autoregressive processes.

Given that o_t and w_t are $I(1)$ series, we assess stationarity in the estimated residuals of the cointegration equation by the unit root test. The tau statistic confirms cointegration ($\tau = -3.64$, less than the critical value -3.37 at the 5 percent level of significance). Note that, because the intercept is insignificant in this equation, a more appropriate critical value would be -2.76 at the 5 percent level of significance, which would increase the evidence of stationary residuals.

However, the evaluation of critical values based on surface models provides a less clear picture. The finite sample MacKinnon (2010, Tables 1 and 2 for the $N = 2$ ‘no trend’ model) has a

TABLE 1 Distribution of occupied (Occ) and job vacancies (Jwc), counts and percentages by Nace (2018Q3)

Nace	Vacancy (Jwc)		Employed (Occ)	
	<i>n</i>	%	<i>n</i>	%
B – Mining and quarrying	311	0.3%	22,500	0.1%
C – Manufacturing	26,449	27.2%	4,213,600	21.5%
D – Electricity, gas, steam and air conditioning supply	372	0.4%	113,000	0.6%
E – Water supply; sewerage, waste management and remediation activities	1018	1.0%	231,700	1.2%
F – Construction	6209	6.4%	1,418,300	7.2%
G – Wholesale and retail trade; repair of motor vehicles and motorcycles	16,050	16.5%	3,159,000	16.1%
H – Transportation and storage	5170	5.3%	1,172,600	6.0%
I – Accommodation and food service activities	9503	9.8%	1,367,000	7.0%
J – Information and communication	6336	6.5%	605,200	3.1%
K – Financial and insurance activities	2776	2.9%	612,400	3.1%
M – Professional, scientific and technical activities	7633	7.9%	1,385,600	7.1%
N – Administrative and support-service activities	6549	6.7%	986,000	5.0%
P – Education	961	1.0%	1,553,400	7.9%
Q – Human health and social work activities	5174	5.3%	1,809,400	9.2%
R – Arts, entertainment and recreation	1073	1.1%	274,900	1.4%
S – Other service activities	1479	1.5%	677,400	3.5%
TOTAL NACE activities	97,063	100.0%	19,602,000	100.0%

critical value equal to -3.580 at the 5 percent level of significance (-3.337 , the asymptotic CV), whereas it is equal to -4.350 at the 1 percent level of significance (-3.900 the asymptotic critical value). Hence, cointegration is not refused at the 5 percent level of significance but is refused at the 1 percent level.

With this note of caution in mind, the estimated cointegrating relationship ($\alpha_t = 0.168(0.185) + 0.038(0.007)w_t$, with standard errors in parenthesis) shows the w_t coefficient is significant (at the 1 percent level), meaning that there is a structural and non-spurious long-run relationship between the two series. When seasonal returns for vacancies increase by 10 percent from $t-4$ to t , the seasonal returns for numbers employed in the same time window increase by almost 0.4 percent (long-run coefficient).

However, seasonal changes in numbers employed may not respond by this full amount within the same quarters of consecutive years. The VECM, taking short-term adjustments into account, reveals a more complete picture. The estimated VECM is shown in Eqs (1) and (2):

$$\Delta o_t = \frac{0.181}{(0.118)} - \frac{0.555e_{t-1}}{(0.127)***} + \frac{0.072\Delta o_{t-1}}{(0.175)} - \frac{0.028\Delta w_{t-1}}{(0.010)*} \quad (1)$$

$$\Delta w_t = \frac{2.879}{(2.380)} + \frac{0.477e_{t-1}}{(2.909)} - \frac{1.049\Delta o_{t-1}}{(3.628)} - \frac{0.006\Delta w_{t-1}}{(0.211)} \quad (2)$$

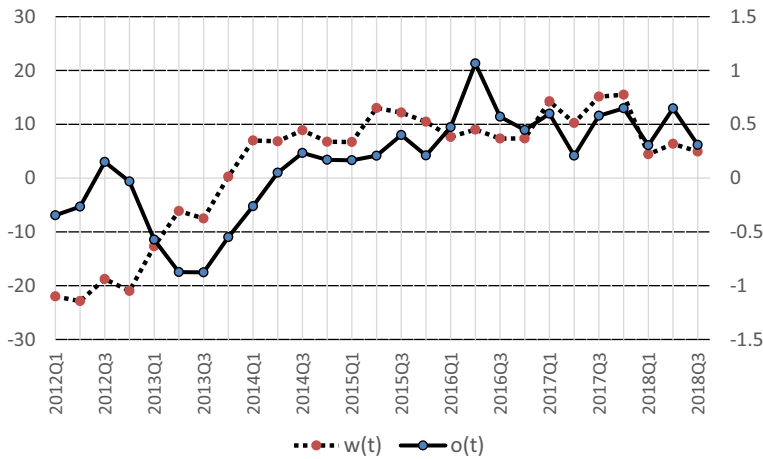


FIGURE 2 Quarterly seasonal returns of occupied $o(t)$ and job vacancies $w(t)$, 2012Q1 – 2018Q3.

In the equation for the employed series, the coefficient of the error correction term (ECT) ($e_{t-1} = o_{t-1} - 0.038w_{t-1}$) is strongly significant (at the 1 percent level) and, in a weaker fashion, the ECT of Δw_{t-1} is also significant (at the 10 percent level). The ECT indicates that Δo_t (short-term changes compared with the same changes one year before) falls when there is a positive cointegrating error ($o_{t-1} > 0.038w_{t-1}$) and that the adjustment of o_t will be around 56 percent of the deviation of o_{t-1} from its long-run equilibrium $0.038w_{t-1}$.

It is worth noting the non-significance of all parameters in Eq. (2), indicating that Δw_t does not react either to a cointegrating error or to past values of Δo_t (thus exhibiting exogeneity) and to past values of Δw_t (confirming the structure found in the auto.arima). The validity of the estimated models in Eqs (1) and (2) is proven by tests applied to the residuals (normally distributed and homoscedastic with no serial correlation).

Hence, the overall finding is that if we fix the level of significance of cointegration at the 5 percent level, there is both a long-run (contemporaneous between levels) and a short-term (between lags of differences) relation among the series. Furthermore, the seasonal returns for numbers employed react to the seasonal returns for vacancies, but not vice versa. This implies Granger causation from movement in vacancies to movements in numbers employed.

The data demonstrate that when demand for labour is strong, the level of vacancies will generally rise, which tends to lead to higher growth in employment as these vacancies are filled. This may occur during a period of economic recovery when the vacancy rate increases because firms post more job openings. This is, in turn, associated with lower unemployment rates and higher employment. This is confirmed by a recent study by the European Central Bank (Consolo and da Silva, 2019) concerning the conditions of the European labour market as described by the Beveridge curve during the period 2004–2019. The study demonstrates that after the EU crisis from 2012–2014 (demonstrated by an outward shift of the Beveridge curve with a decrease in vacancies and a steady increase in unemployment), the curve for Italy, although significantly below the previous peaks in Q1 2011 and Q2 2007, steepens significantly from 2016. It shows a moderate recovery phase with a typical anticlockwise movement, characterized by an increase in vacancies that is faster than the decrease in unemployment.

The main cause for this recovery appears to be an increase in the job-finding rate (the fraction of those unemployed that move out of unemployment) and a decrease in the separation

TABLE 2 Not cointegrated sectors: ARIMA models, I order, model and Granger causality

Nace (Sector)	Endo-genous	Arima model	I order	Model	Granger (p-value)
N – Administrative and support-service activities	o_t	ARIMA(1,1,0) (2,0,0) ₄	I(1)	VECM	0.234
	w_t	ARIMA(0,1,0)	I(1)		0.755
R – Arts, entertainment and recreation	o_t	ARIMA(0,0,2) (0,0,1) ₄	I(0)	VAR in levels	0.055
	w_t	ARIMA(1,0,0)	I(0)		0.497
I – Accommodation and food service activities	o_t	ARIMA(1,1,0) (0,0,1) ₄	I(1)	No cointegration VAR in 1st diff	0.239
	w_t	ARIMA(0,1,0) (2,0,0) ₄	I(1)		0.745
G – Wholesale and retail trade; repair motor vehicles and motors	o_t	ARIMA(1,0,0)	I(0)	VAR in levels	0.531
	w_t	ARIMA(0,0,0)	I(0)		0.312
Q – Human health and social work activities	o_t	ARIMA(0,0,0) (0,0,1) ₄	I(0)	VAR in levels	0.365
	w_t	ARIMA(0,0,0) (0,0,1) ₄	I(0)		0.663

rate (the likelihood that an employed person leaves or loses their current job). By contrast, labour market efficiency (measured by the number of people finding jobs given a certain vacancy-unemployment ratio) during 2006–2019 has not fully recovered to pre-crisis levels (Consolo and da Silva 2019).

4.2 | Modelling sectors

The previous analysis is repeated for most representative sectors (by count), omitting those sectors representing less than 5 percent of the overall vacancies and numbers employed in the analysis period. The order of integration (and cointegration) is assessed using both classical tau statistics and MacKinnon's (2010) finite-sample critical values at the 5 percent and 1 percent level (using models with different deterministic components, depending on their significance). The principal results are organized into separate sectors, those (in Table 2) that are not cointegrated (for various reasons, including stationarity, a different order of integration or an I(1) series with a nonsignificant cointegration coefficient) and those sectors that are cointegrated (Table 3).

In Table 2, we also include the sector N administrative and support-service activities. Despite formally finding cointegration ($\tau = -3.663$, at the 5 percent level of significance), neither the ECT coefficient nor the coefficients of the lagged exogenous variables are significant in the VECM. For a better picture, Figures 3–5 show the seasonal returns and first differences for the employment and vacancy series for the cointegrated (at the 1 and 5 percent level of significance, using MacKinnon critical values) and not cointegrated sectors, respectively.

Table 2 provides the best ARIMA representation, the order of integration, the model used (with the same order of integration) and the significance of the lagged values of the exogenous variable for each endogenous variable (Granger column), which is useful for establishing

TABLE 3 Cointegrated sectors: ARIMA models, I order, tau statistic on residuals of the cointegration

Nace (Sector)	Endo-genous	Arima model	I order	tau on EG residuals
Cointegration at the 1% level				
C – Manufacturing	o_t	ARIMA(0,1,3)	I(1)	$t = -4.750$
	w_t	ARIMA(0,1,0)(1,0,0) ₄	I(1)	
J – Information and communication	o_t	ARIMA(0,1,1)(0,0,1) ₄	I(1)	$t = -5.190$
	w_t	ARIMA(0,1,0)(0,0,1) ₄	I(1)	
Cointegration at the 5% level				
M – Professional, scientific and technical activities	o_t	ARIMA(1,1,0)	I(1)	$t = -4.088$
	w_t	ARIMA(0,1,0)(1,0,0) ₄	I(1)	
F – Construction	o_t	ARIMA(2,1,0)(0,0,1) ₄	I(1)	$t = -3.848$
	w_t	ARIMA(1,1,0)(2,0,1) ₄	I(1)	
H – Transportation and storage	o_t	ARIMA(0,1,2)	I(1)	$t = -3.262$
	w_t	ARIMA(1,1,0)(0,0,1) ₄	I(1)	
All sectors	o_t	ARIMA(2,1,0)(1,0,0) ₄	I(1)	$t = -3.635$
	w_t	ARIMA(0,1,0)(0,0,1) ₄	I(1)	

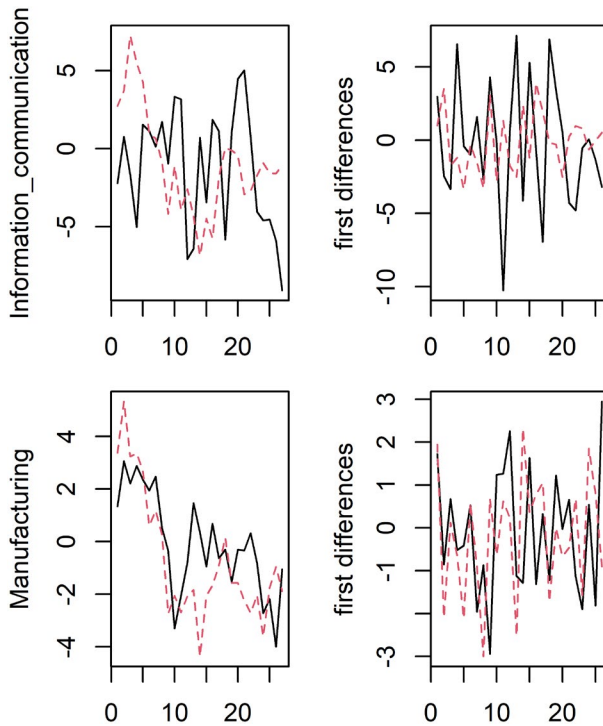


FIGURE 3 Seasonal returns and 1st differences of occupied and vacancy (dotted), cointegrated sectors at the 1 percent level.

Granger causality. For accommodation and food service activities (I), the series are not cointegrated, whereas for three sectors (R, G and Q), there are flat seasonal returns (stationarity) in the original series.

Examining the parameter significance of the lagged exogenous variable in each equation for these sectors, the lagged values of vacancy returns do not appear to be useful for forecasting future values of returns for employment and vice versa. This implies the absence of any significant dynamic relationship between the series for these sectors. This may occur for several reasons. For example, given that employment in some sectors has remained stable over time, an increase in the stock of vacancies over the same time period might indicate a growing unmet demand, unfilled vacancies or an increase in the rate of turnover.

Table 3, which includes the cointegrated sectors, adds diagnostic data on cointegration (tau statistics on the residuals of the EG first step) and separate cointegrated sectors among those that are cointegrated at the 1 and 5 percent level of significance, based on the critical values of MacKinnon (2010). Specifically, the finite sample MacKinnon (2010) critical value is equal to -3.580 at the 5 percent level of significance (-3.337 , the asymptotic CV), whereas this is equal to -4.350 at the 1 percent level of significance (-3.900 , the asymptotic critical value). For these sectors, Table 4 sets out the main results for the coefficients of the cointegrating relations, the error correction coefficient in the VECM (for the employed equation, only) and the goodness of fit statistics for both estimated models.

A large portion of the overall long-run equilibrium among series depends on the manufacturing sector, which shows the strongest relation among the sectors in terms of both the estimated coefficients and significance. When the seasonal returns for vacancies in manufacturing increase

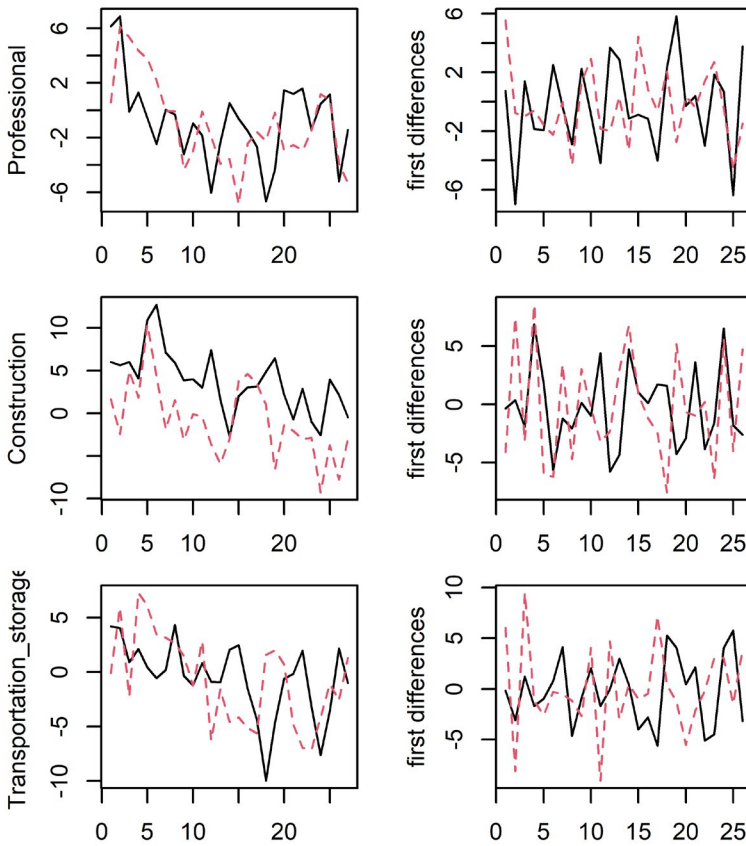


FIGURE 4 Seasonal returns and Ist differences of occupied and vacancy (dotted), cointegrated sectors at the 5 percent level.

by 10 percent (the percentage increase in vacancies between the corresponding quarters of two consecutive years), the seasonal returns for numbers employed increase by 0.54 percent (more than the overall pattern). We find a similar pattern for the construction (0.44 percent), professional (0.39 percent) and ICT (0.34 percent) sectors (although the evidence of cointegration is less strong for the first two sectors than the ICT sector). The transportation and storage sector demonstrates a weaker long-run relationship.

Regarding the adjustment of occupied to vacant posts in the short-run (Δo_t), the construction and ICT sectors have the fastest ECT. This means that the adjustment of occupied-post returns for consecutive quarters, seasonally adjusted (Δo_t), will be about 91 percent and 83 percent, respectively of the deviation from its long-run equilibrium achieved one-quarter before (o_{t-1}) for vacancy returns (values that are faster than other sectors). By contrast, for manufacturing, the adjustment is slower (65 percent).

By examining the correlation between the job vacancy and employment series, the strength of the relationship between the two series at different lags is assessed as a further robustness check (see Table 5).

The correlation structure confirms the cointegration results; the relations among series are weak for sectors that are not cointegrated. In contrast, for cointegrated sectors, the correlations with employment growth one and two quarters ahead are higher than the correlations with employment growth one-quarter earlier. Apart from contemporaneous correlation, the strongest

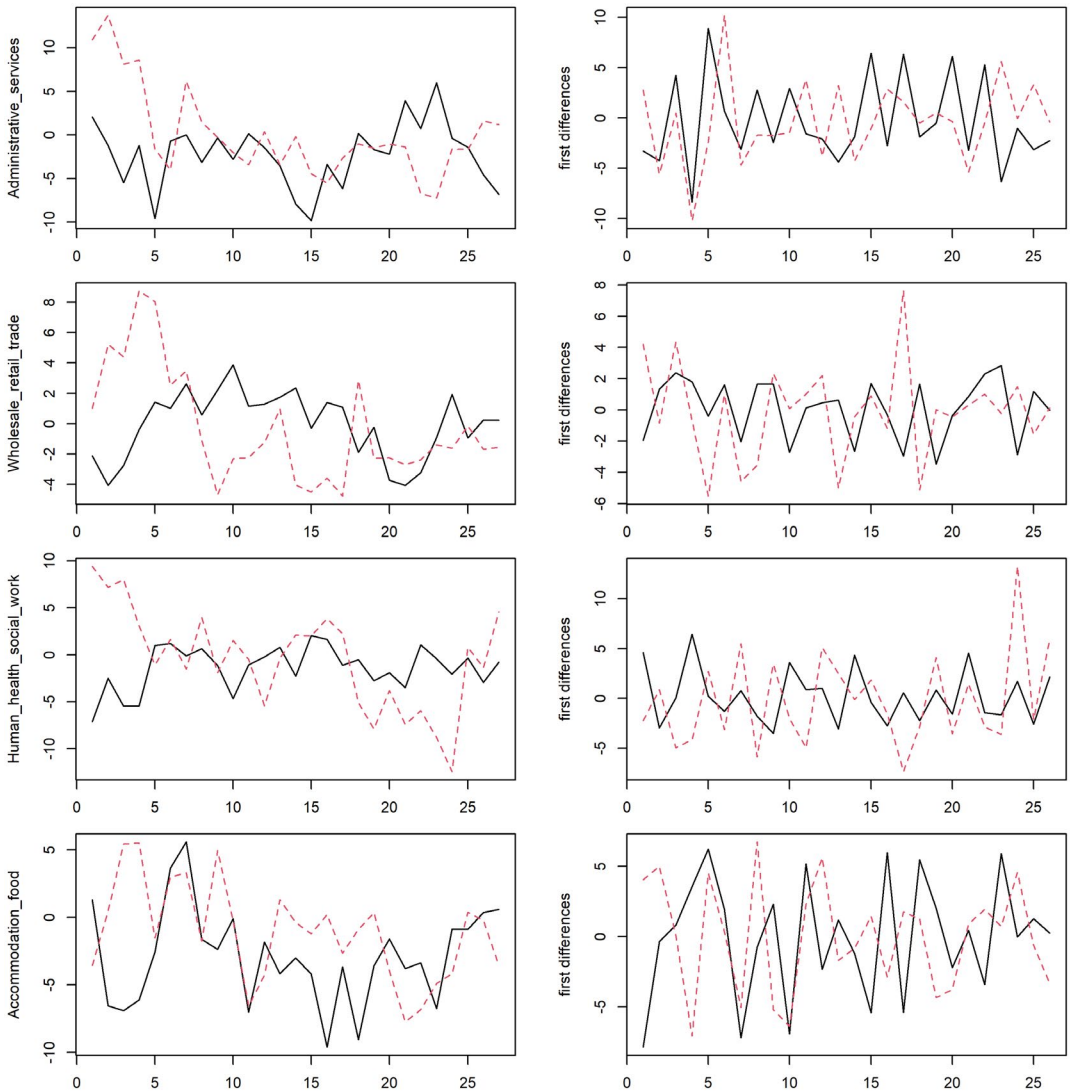


FIGURE 5 Seasonal returns and Ist differences of occupied and vacancy (dotted), not cointegrated sectors.

correlation coefficients are found between the job vacancy series in one-quarter and the employment series in the subsequent quarter. This confirms that growth in job vacancies leads employment growth by about one-quarter.

5 | CONCLUSIONS

The overall view, considering the small time series, is that there is a structural relationship between the series. As regards the direction of causation, the results imply Granger causation from vacancies to employment (as seasonal returns for occupied posts react to seasonal returns for vacancies) but not vice versa, indicating that the (seasonal) vacancy growth is a leading indicator

TABLE 4 Cointegration relations (normalized on o_t) and VECM results (o_t as endogenous variable), by sectors

Nace (Sector)	Endo-genous	Cointegration relations		VECM model	
		Long-run coefficient	R-square	Error correction coefficient	R-square
Cointegration at the 1% level					
C – Manufacturing	o_t	0.054(0.011)***	0.486	–0.654(0.226)***	0.324
J – Information and communication	o_t	0.034(0.012)*	0.237	–0.8301(0.197)***	0.464
Cointegration at the 5% level					
M – Professional, scientific and technical activities	o_t	0.039(0.017)*	0.178	–0.806(0.232)**	0.497
F – Construction	o_t	0.044(0.0139)**	0.290	–0.907(0.238)**	0.436
H – Transportation and storage	o_t	0.026(0.016)*	0.298	–0.817(0.182)***	0.464
All Sectors	o_t	0.038(0.007)***	0.308	–0.555(0.127)***	0.454

of (seasonal) employment growth and that seasonal movements for vacancies are weakly exogenous. Specifically, quarterly adjustment of seasonal variation in occupied posts will be about 55 percent of the deviation of o_{t-1} from its long-run equilibrium with w_{t-1} . In terms of sectors, manufacturing and ICT show the strongest cointegrating relation among the series (although for ICT, the contemporaneous relation among seasonal returns is weak in quality). The construction sector demonstrates the fastest short time adjustments (for seasonal returns) for consecutive quarters.

Our results suggest that as well as being a source of information on current labour market conditions, vacancies provide leading information on employment growth, both contemporaneously and one-quarter ahead. The forecasting performance of models including these indicators will provide a better assessment of employment developments for specific sectors. In fact, because this relationship also holds when the series are disaggregated at the sectoral level, vacancies can be opportunely employed as a leading indicator of labour market trends. Giving priority to the strength of cointegration over the significance of coefficients and R-squared in the cointegration equation, this appears particularly true for the manufacturing and ICT sectors.

From an economic perspective, despite deteriorating labour market efficiency for the period 2006–2019 as compared with pre-crisis times, and as previously reported for aggregate sectors in Italy, both the manufacturing and ICT sectors show that a strong labour demand increases the level of vacancies; this, in turn, tends to lead to higher employment growth. One possible interpretation is that the matching rates (the likelihood that a vacancy is filled) at given separation rates are higher for these than for other sectors. In fact, ICT and manufacturing generally involve specialized and technical employees who are better able and more willing to move across

TABLE 5 Cross-Correlations among seasonal returns $o(t)$ and $w(t \pm k)$

	$w(t-2)$	$w(t-1)$	$w(t)$	$w(t+1)$	$w(t+2)$
Cointegrated sectors					
C – Manufacturing	0.613	0.706	0.697	0.244	0.180
J – Information and communication	0.228	0.488	0.487	0.400	0.290
M – Professional, scientific and technical activities	0.107	0.531	0.422	0.108	0.241
F – Construction	0.568	0.611	0.539	0.144	0.172
H – Transportation and storage	0.276	0.389	0.547	0.027	0.058
Not cointegrated sectors					
N – Administrative and support-service activities	-0.116	-0.303	0.032	0.131	-0.021
R – Arts, entertainment and recreation	-0.522	-0.334	0.002	0.072	0.209
I – Accommodation and food service activities	0.067	0.190	0.063	-0.007	0.091
G – Wholesale and retail trade; repair motor vehicles and motors	0.116	0.046	-0.178	-0.276	-0.395
Q – Human health and social work activities	-0.063	-0.208	-0.245	-0.197	-0.087

different industries or geographies. Moreover, higher matching rates may result from developments in recruitment technology, such as online job advertisements, that particularly focus on professional figures of such industries (Garasto et al., 2021).

However, our paper has some limitations. Notwithstanding these promising results, tracking the evolution of the relationship between employment and vacancies over short time periods is not sufficient to derive conclusions regarding possible shifts in the labour supply in response to labour demand that points to improving or worsening conditions in certain sectors. In fact, the econometric results regarding stationarity, cointegration and estimation of VAR and VECM models are based on a small sample with a maximum of 27 observations; such a small sample may yield misleading test statistics. The brevity of the available series is thus an obstacle to a deeper and more accurate validation analysis of the estimated models by, for example separating the data into a training set (for modelling relationships) and a test set (for evaluating the stability and significance of estimated parameters on an independent set of data). Nevertheless, in this study, we use various diagnostics (robust critical values, cross-correlations and Arima representations) as validation checks. These essentially, albeit not ubiquitously, confirm the inferential results for the integration order and cointegration analysis based on traditional critical values. In future research, longer time series should be used to address these important issues.

The absence of vacancies for firms with fewer than 10 employees may cause severe bias in the stock measurements. However, another official Italian survey, the Excelsior survey, conducted by the Italian Union of the Chambers of Commerce (Unioncamere) and funded by the Ministry of Labour, can be used as a benchmark and provides detailed information on the characteristics of labour demand in the country. The most recent Excelsior survey (for 2018) estimates that only 9.8 percent of firms with nine employees or fewer had used specialized agencies, national job centres or information channels (such as web job postings, social networks and specialized journals) as their main job search channel in the previous 12 months (Excelsior, 2019). Moreover, because official data are generally implemented through quarterly surveys and offer very little detail at the occupational level, they can provide only a limited assessment of the true underlying labour market conditions in real-time.

Policy intervention relies on having good information on current and future skills needs. Such information facilitates examination of the changing patterns of skills demand and allows skills imbalances to be addressed. This involves a detailed examination of sectoral as well as occupational employment changes and the implications of these at both the micro and macro levels. To this end, web job portals may be useful auxiliary sources for data on labour demand (see Papoutsoglou et al., 2019, for a recent review; see also Cedefop, 2017); however, using web (big) data in the form of online job advertisements (OJA) is a complex undertaking. The use of such data, in fact, requires the resolution of problems linked to definition and data collection (duplication, selection of stable and credible sources, removing non-work ads, selection of active ads etc.), the transformation from OJA flows to stocks (Garasto et al., 2021; Turrell et al., 2019) and issues such as non-representativeness and lack of coverage at the population level (Couper, 2013; Fan et al., 2014; Japac et al., 2015; Kureková et al., 2015; Tam & Clarke, 2015).

A recent institutional research project Cedefop (2020), coordinated by Bicocca University (Crisp Research Centre) and funded by Cedefop and Eurostat, focuses on these issues for providing timely and relevant statistics for the European labour market and associated skills' demands using all available data sources, including emerging digital sources and especially OJAs.

Resolving these issues would allow for official statistics that provide measurements at lower cost and higher frequency and that integrate big data and traditional surveys. This methodological

agenda, with the aim of developing better statistical tools for ensuring the validity of web data, constitutes one of the main challenges for future labour market analyses.

CONFLICTS OF INTEREST

None declared.

ACKNOWLEDGEMENTS

Open Access Funding provided by Universita degli Studi di Milano-Bicocca within the CRUI-CARE Agreement.

ENDNOTES

¹ Various measurement issues associated with vacancy data make it difficult to interpret the recorded point-in-time stock of vacancies in any given quarter. The point-in-time vacancy stock does not necessarily represent all the vacancies that opened during the period, due to the possible unmeasured flow of new vacancies posted and filled shortly before or after that point.

² https://ec.europa.eu/eurostat/cache/metadata/EN/jvs_esqrs_it.htm.

REFERENCES

- Amoah, B. (2000) Help-wanted index. *Perspectives on Labour and Income. Statistics Canada*, 75–001-XPE, 14–18.
- Askatas, N. & Zimmermann, K. F. (2009) Google econometrics and unemployment forecasting. *IZA Discussion Paper*, 4201.
- Australian Bureau of Statistics. (2003) *Do Job Vacancies Provide a Leading Indicator of Employment Growth?* Australian Labour Market Statistics. Catalogue no. 6105.0, April 2003.
- Australian Government (2019) *Department of Employment, Skills, Small and Family Business, Annual Report 2018–19*. AU: Australian Government Press. Available at: <http://lmip.gov.au/default.aspx?LMIP/GainInsights/EmploymentProjections>.
- Banerjee, A., Dolado, J.J., Galbraith, W. & Hendry, D. F. (1993) *Co-integration, Error Correction, the Econometric Analysis of Non-stationary Data*. Oxford: Oxford University Press.
- Banerjee, A., Dolado, J. J. & Mestre, R. (1998) Error-correction mechanism tests for cointegration a single-equation framework. *Journal of Time Series Analysis*, 19, 267–283.
- Blanchard, O. J., Diamond, P., Hall, R. E. & Yellen, J. (1989) The Beveridge curve. *Brookings Papers on Economic Activity*, 1, 1–60.
- Cedefop (2017). *Big data analysis from job vacancies project*. Available at: <http://www.cedefop.europa.eu/en/event-s-and-projects/projects/big-data-analysis-online-vacancies>
- Cedefop (2020) *Towards the European Web Intelligence Hub – European system for collection and analysis of online job advertisement data (WIH-OJA): OF2-Cedefop: representativeness of OJAs*. Available at: <https://www.cedefop.europa.eu/da/about-cedefop/public-procurement/towards-european-web-intelligence-hub-european-system-collection-and-analysis-online-job>
- Cheung, Y. W. & Lai, K. S. (1993) Finite sample sizes of Johansen's likelihood ratio tests for cointegration. *Oxford Bulletin of Economics and Statistics*, 55, 313–328.
- Consolo, A. & da Silva, A. (2019) The euro area labour market through the lens of the Beveridge curve. *ECB Economic Bulletin*, 4/2019, European Central Bank.
- Couper, M. P. (2013) Is the sky falling? New technology, changing media, and the future of surveys. *Survey Research Methods*, 7, 145–156.
- D'Amuri, F. & Marcucci, J. (2009) Google it! Forecasting the US unemployment rate with a Google job search index. *ISER Working Paper Series 2009-32*, Institute for Social and Economic Research.
- Doornik, J. A. (1998) Approximations to the asymptotic distributions of cointegration tests. *Journal of Economic Surveys*, 12, 573–593.

- Dur, R. A. J. (1999) Mismatch between unemployment and vacancies in the Dutch labour market. *Applied Economics*, 31, 237–244.
- Elsby, M., Michaels, R. & Ratner, D. (2015) The Beveridge curve: a survey. *Journal of Economic Literature*, 53, 571–630.
- Enders, W. (2010) *Applied econometric time series*, 3rd edn. New York: John Wiley & Sons.
- Engle, R. F. & Granger, C. W. J. (1987) Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55, 251–276.
- Engle, R. F. & Yoo, B. S. (1987) Forecasting and testing in co-integrated systems. *Journal of Econometrics*, 35, 143–159.
- Ericsson, N. R. & MacKinnon, J. G. (2002) Distributions of error correction tests for cointegration. *Econometrics Journal*, 5, 285–318.
- European Central Bank (2002) *Monthly bulletin*. Frankfurt: European Central Bank.
- Eurostat. (2018). *Job vacancy rate 2018*. Available at: <https://ec.europa.eu/eurostat/web/products-datasets/-/teilm310>
- Eurostat (2019) *Job vacancy and unemployment rates – Beveridge curve*. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php/Job_vacancy_and_unemployment_rates_-_Beveridge_curve
- Excelsior (2019) *La Domanda di Professioni e di Formazione delle Imprese Italiane nel 2018*. Rome: Unioncamere Press.
- Fan, J., Han, F. & Liu, H. (2014) Challenges of big data analysis. *National Science Review*, 1, 293–314.
- Fondeur, Y. & Karamé, F. (2013) Can Google data help now or forecasting French unemployment? *Econometric Modelling*, 30, 117–125.
- Garasto, S., Djumalieva, J., Kanders, K., Wilcock, R. & Sleeman, C. (2021). *Developing experimental estimates of regional skill demand (ESCoE DP 2021-02)*. ESCoE Discussion Paper 2021-02, 8 March 2021.
- Gorter, C. & van Ours, J. (1994) Matching unemployment and vacancies in regional labor markets: an empirical analysis for the Netherlands. *Papers in Regional Science*, 73, 153–167.
- Gregory, A. W. (1994) Testing for cointegration in linear quadratic models. *Journal of Business and Economic Statistics*, 12, 347–360.
- Haggard-Guénette, C. (1989) Jobs ads: A leading indicator? *Perspectives on Labour and Income. Statistics Canada*, 1, 75–001-XPE, 1–12.
- Harbo, I., Johansen, S., Nielsen, B. & Rahbek, A. (1998) Asymptotic inference on cointegrating rank in partial systems. *Journal of Business and Economic Statistics*, 16, 388–399.
- Hurvich, C. M. & Tsai, C. L. (1989) Regression and time series model selection in small samples. *Biometrika*, 76, 297–307.
- Hyndman, R. J. & Athanasopoulos, G. (2017) *Forecasting: principles and practice*, 2nd edn. OTexts.org.
- Hyndman, R. J. & Khandakar, Y. (2008) Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27, 1–22.
- ILO (2017) *Labour Force Estimates and Projections: 1990–2030*. Geneva: International Labour Office Press.
- Istat (2018) *Quarterly survey on job vacancies and hours worked*. Rome: Istat Press. Available at: <http://siqual.istat.it/SIQual/visualizza.do?id=7779928>
- Jackman, R., Layard, R. & Pissarides, C. (1989) On vacancies. *Oxford Bulletin of Economics and Statistics*, 51, 377–394.
- Jackman, R. & Roper, S. (1987) Structural unemployment. *Oxford Bulletin of Economics and Statistics*, 49, 9–36.
- Japac, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., Lane, J., O’Neil, C. & Usher, A. (2015) *Report on Big data*. American Association for Public Opinion Research (AAPOR). Available at: <http://www.aapor.org/Education-Resources/Reports/Big-Data.aspx>
- Johansen, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12, 231–254.
- Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59, 1551–1580.
- Johansen, S. (1995) *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: Oxford University Press.
- Johansen, S. & Juselius, K. (1990) Maximum likelihood estimation and inference on cointegration – with applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52, 169–210.
- Kureková, L. M., Beblavý, M. & Thum-Thyssen, A. (2015) Using online vacancies and web surveys to analyse the labor market: a methodological inquiry. *IZA Journal of Labor Economics*, 4, 1–18.
- MacKinnon, J. G. (1991) Critical values for cointegration tests. In: Engle, R. F. & Granger, C. W. J. (Eds.) *Long-run economic relationships: readings in cointegration*. Oxford: Oxford University Press, pp. 267–276.

- MacKinnon, J. G. (1996) Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics*, 11, 601–618.
- MacKinnon, J.G. (2010) Critical values for cointegration tests. *Queen's Economics Department Working Paper*, 1227, Queen's University, Department of Economics, Kingston (Ontario)
- MacKinnon, J. G., Haug, A. A. & Michelis, L. (1999) Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of Applied Econometrics*, 14, 563–577.
- Mandrone, E., Tancioni, M. & Laj, S. (2010) Vacancies as employment predictor: dynamic properties of the series. In: Mandrone, E. (Ed.) *Job vacancies in the Italian labour market. The new ISFOL "Help Wanted" Time Series*. Rubettino, Roma: Isfol "TEMI E RICERCHE".
- OECD (2018) *Approaches to anticipating skills for the future of work*. Geneva: OECD Press.
- Osterwald-Lenum, M. (1992) A note with quantiles of the asymptotic distribution of the maximum hood cointegration rank test statistics. *Oxford Bulletin of Economics and Statistics*, 54, 461–472.
- Papoutsoglou, M., Ampatzoglou, A., Mittas, N. & Angelis, L. (2019) Extracting knowledge from online sources for software engineering labor market: a mapping study. *IEEE*, 7, 157595–157613.
- Pesaran, M. H., Shin, R. J. & Smith, R. J. (2000) Structural analysis of vector error correction models exogenous I(1) variables. *Journal of Econometrics*, 97, 293–343.
- Phillips, P. C. B. & Ouliaris, S. (1990) Asymptotic properties of residual based tests for cointegration. *Econometrica*, 58, 165–193.
- Reinsel, G. C. & Ahn, S. K. (1988) *Asymptotic properties of the likelihood ratio test for cointegration in the non-stationary vector AR model*. Technical Report, Department of Statistics, University of Wisconsin, Madison.
- Ruth, F. J., Schouten, J. G. & Wekker, R. F. (2006) Statistics Netherlands business cycle tracer. Methodological aspects; concept, cycle computation and indicator selection. *Statistics Netherlands Discussion Paper*, 5, 1–64.
- Said, S. E. & Dickey, D. A. (1984) Testing for unit roots in Autoregressive-Moving Average Models of unknown order. *Biometrika*, 71, 599–607.
- Sims, C. (1980) Macroeconomics and reality. *Econometrica*, 48, 1–48.
- Štefánik, M. (2012) Internet job search data as a possible source of information on skills demand (with results for Slovak University graduates). In: Cedefop (Ed.) *Building on skills forecasts—Comparing methods and applications*. Luxembourg: European Union Press.
- Steinmetz, S., Tjidsens, K. & de Pedraza, P. (2009). WP 76-Comparing different weighting procedures for volunteer Web surveys. AIAS Working Paper 09/76, Universiteit van Amsterdam.
- Stewart, C. (2011) A note on spurious significance in regressions involving I(0) and I(1) variables. *Empirical Economics*, 41, 565–571.
- Tam, S.M. & Clarke, F. (2015) Big data, statistical inference and official Statistics, *Methodology Research Papers*, 1351.0.55.054, Australian Bureau of Statistics, Canberra, CA.
- Turrell, A., Speigner, B.J., Djumalieva, J., Copple, D. & Thurgood, J. (2019). Transforming naturally occurring text data into economic statistics: The case of online job vacancy postings. 2019. Technical report.
- U.S. Bureau of Labor Statistics (2019) *Employment Projections*. Available at: <https://www.bls.gov/emp/documentation/projections-methods.htm#labor>
- Valletta, R. (2005) Help-wanted advertising and job vacancies. *Federal Reserve Bank of San Francisco Economic Letter*, 2, 1–3.
- Wilson, R., Sofroniou, N., Beaven, R., May-Gillings, M., Perkins, S., Lee, M., Glover, P., Limmer, H. & Leach, A. (2016) *Working Futures 2014-2024*. Technical report on sources and Methods. Cambridge, UK: Cambridge Econometrics.
- Zagorsky, J. L. (1998) Job vacancies in the United States: 1923 to 1994. *Review of Economics and Statistics*, 80, 338–345.

How to cite this article: Lovaglio, P.G. (2022) Do job vacancies variations anticipate employment variations by sector? Some preliminary evidence from Italy. *Labour*, 36, 71–93. <https://doi.org/10.1111/labr.12213>

APPENDIX 1

Choosing the correct strategy – VAR (relationship among stationary variables) or VECM (relationship among non-stationary but cointegrated series) – requires evaluation of the stationarity of the o_t and w_t series and order of integration d . Given a non-stationary variable x with a stochastic trend, d indicates the number of times that x must be differenced in order to be made stationary, briefly $x \sim I(d)$. Common situations, especially for seasonally adjusted series, are $I(0)$, stationary variables, and $I(1)$, first difference stationary variables $\Delta x_t = x_t - x_{t-1}$. Other situations of non-stationarity may depend on a deterministic linear or squared trend (no unit root).

VAR

If both series o_t and w_t were stationary, a VAR model was specified. A VAR(p) is a general framework that describes a system in which each variable is a function of its own p lagged values and the lag of the other variable in the system. Therefore, it uses both variables as endogenous (o_t is a function of o_{t-1}, \dots, o_{t-p} and w_{t-1}, \dots, w_{t-p} , whereas w_t is a function of its own p lags and those of o_t).

Typically, when both trend and seasonality are present, we may need to apply a non-seasonal first difference to seasonal differenced counts: $\Delta o_t = o_t - o_{t-1} = \log(\text{Occ}_t/\text{Occ}_{t-4}) - \log(\text{Occ}_{t-1}/\text{Occ}_{t-5})$, measuring the differences between consecutive seasonal returns. The same holds for Δw_t . As Δo_t can be written as $\Delta o_t = \log(\text{Occ}_t) - \log(\text{Occ}_{t-1}) - [\log(\text{Occ}_{t-4}) - \log(\text{Occ}_{t-5})]$, this is the amount by which the change from the previous period to the current period is different from the change that was observed exactly one year earlier. In other words, the difference in the quarter-to-quarter seasonally adjusted estimates is a direct measure of the change in the number of people working, after taking into account the changes expected because of the variation in seasonal employment between these two quarters.

Hence, if o_t and w_t are $I(1)$ series, a bivariate VAR(p) can be specified (as shown in Eqs A1.1 and A1.2, for $P = 1$), for first differenced seasonally adjusted returns Δo_t and Δw_t :

$$\Delta o_t = \pi_{10} + \pi_1 \Delta o_{t-1} + \gamma_{11} \Delta w_{t-1} + \epsilon_{ot} \quad (\text{A1.1})$$

$$\Delta w_t = \pi_{20} + \pi_2 \Delta w_{t-1} + \gamma_{21} \Delta o_{t-1} + \epsilon_{wt} \quad (\text{A1.2})$$

where π and γ are structural parameters, and ϵ_t is stochastic white noise error, possibly intercorrelated. The significance of structural parameters demonstrates possible relationships among the series. For example, Eq. A1.1 indicates whether growth in seasonal returns for employed is significantly related to its own past value (π_1) and/or to the lagged growth in vacancies (γ_{11}).

VECM

Relationships in a VAR, relating changes in non-stationary variables (as specified in Eqs A1.1 and A1.2), are short term and cannot establish long-run relationships between the levels of both variables.

As an exception, the presence of cointegration, where two or more series can be individually $I(1)$ non-stationary but some linear combination of them is stationary, allows the long-run relation of time series to be assessed even when series are not stationary but $I(1)$ or first difference (stochastic trend) stationary.

To assess cointegration among our series, we used the Engle and Granger (1987) two-step strategy (EG). First, we fitted a linear regression $o_t = \beta_1 + \beta_2 w_t + \epsilon_t$ (with intercept, since here it

has economic meaning). Second, we checked whether estimated least-square residuals $e_t = o_t - b_1 - b_2 w_t$ were stationary using ADF tau statistics. Here, we normalized on o_t (i.e. as a dependent variable in the linear regression), simply because it is more familiar to think of coefficients in terms of the (seasonal) percentage change in employed than in terms of changes in vacancies. This does not, however, preclude the hypothesis that both series co-move in a symmetric causal relation.

In the presence of cointegrated I(1) variables, VECM(p) is a very popular model allowing the specification of a long-run relationship between series (the cointegration equation, among variables in levels), as well as for short-run relationships (i.e. changes of non-stationary variables, measured in differences, to changes in the cointegrating term and lags of endogenous and exogenous variables), towards the long-run relationship.

An example of VECM(1) is presented in Eqs A2.1 and A2.2:

$$\Delta o_t = \pi_{10} + \alpha_1 e_{t-1} + \pi_1 \Delta o_{t-1} + \gamma_{11} \Delta w_{t-1} + \epsilon_{ot} \quad (\text{A2.1})$$

$$\Delta w_t = \pi_{20} + \alpha_2 e_{t-1} + \pi_2 \Delta w_{t-1} + \gamma_{21} \Delta o_{t-1} + \epsilon_{wt} \quad (\text{A2.2})$$

where e_{t-1} are the lagged values of the residual in the cointegration relationship (estimated as in EG) known as the error correction term (ECT), which expresses the long-run relationship among the o_t and w_t series (i.e. their stationary linear combination) and ϵ_{ot} , ϵ_{wt} are stationary white noise terms, possibly correlated.

Coefficients α_1 and α_2 are the error correction coefficients (so called because they show how much Δo_t and Δw_t respond to the past cointegrating error e_{t-1}). The value of α_1 determines the speed of adjustment between consecutive periods of endogenous series (Δo_t) and should always be negative in sign in order for the system not to diverge from its long-run equilibrium. Comparing Eqs A1.1–A1.2 to A2.1–A2.2, VECM is a special form of VAR for I(1) variables that are cointegrated.

In a VECM(p) with cointegration, there should be a causation mechanism, defined as Granger causality, in at least one direction; coefficients of past values of w_t are significant for forecasting o_t values, or vice versa, or both. With larger p , these joint hypotheses can be tested using the classical F-test, whereas for $p = 1$ (as in Eqs A2.1 and A2.2), the Granger causality simplifies to assess the significance of lagged exogenous terms in both equations (gamma parameters). Furthermore, a pure VECM(p), even in the extreme case of a nonsignificant ECT parameter and thus coinciding with a VAR(p) and necessarily involving a Granger causation mechanism, is typically better (fit) than a pure autoregressive representation of the analysed series. To this end, the joint test on zero gamma coefficients in a VAR equation ($\gamma_{11}, \dots, \gamma_{1p}$, e.g. extending until p lags Eq. A1.1) can be seen as a test between two (nested) VAR(p) and AR(p) models.