UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Measuring Labor Demand using Job Ads

Kevin Dee Boman

Wilhelm Åkesson

Supervisor: Andreas Dzemski

Master's thesis in Economics, 30 hec

Spring 2021

Graduate School, Gothenburg School of Business, Economics and Law, Sweden

Abstract

Using the largest digital job ads platform in Sweden, we construct a daily labor demand index based on millions of job ads. This daily index functions as a less expensive, high frequency complement to official survey data. To measure the precision of this index we compare it to official survey statistics and find that it tracks labor demand. The granularity of the index enables disaggregation by region and occupation on a level of detail not found in conventional methods. It is also used to construct empirical Beveridge curves. This study contributes to a growing scientific literature on using internet data to analyze the labor market.

Keywords: labor market, Beveridge curve, big data, vacancy rate, internet data, time series

Acknowledgements

We are grateful for the invaluable help and professional advice we received from Simon Benjaminsson at Arbetsförmedlingen. We would also like to express our sincerest gratitude to our supervisor Andreas Dzemski at the department of Economics. Without his guidance on structure and econometrics this thesis would not have been possible.

Contents

1 Introduction

Policy change, such as changing the payroll tax, and thus affecting the price of labor could have implications for the demand of labor. To evaluate effects of policy change and employment programs, it is important that policy makers have an accurate and timely measure of labor demand. We use data from the Swedish official job ads platform, Arbetsförmedlingen, to compute a new index of labor demand. This index measures labor demand in terms of how many jobs are advertised online for any given day. We show that our index tracks labor demand and can be used as a complement to current methods.

The constructed index has a high level of detail regarding demand among specific occupations and regions. We use this fact to disaggregate the index by occupational group and region, displaying how it can be used for further studies of labor demand. The results are also analyzed through a theoretical modelling of the Beveridge curve that relates vacancies with the unemployment rate for different occupations. This illustrates how the index can be used to examine the Beveridge curve theory empirically. With this study we answer the following research question:

Can an index based on digital job ads be used to measure labor demand?

The main method for measuring labor demand in Sweden today is through Statistics Sweden's quarterly vacancy survey (2020). Sending out quarterly surveys to firms is the main method for measuring labor demand in many developed countries (Lovaglio et al., 2019). The daily index we have constructed can complement these kinds of surveys in several important ways. Given the granularity and frequency of our data - available per day - this index provides information on spikes in labor demand the same day as these occur. This level of detail and frequency can benefit decision makers, as they can be made aware of changes in labor demand much earlier. The daily frequency also adds more observations which enables higher statistical power when analyzing Swedish labor demand. Since the whole process can be automatized the index is also less expensive to create than the quarterly vacancy surveys that decision makers currently rely on.

Internet data is gaining importance as a source for labor market research (Lenaerts et al., 2016). Both firms and employees increasingly look for labor and work respectively online, providing more data on their behavior. A study similar to ours was performed in Italy by Lovaglio et al. in 2019 using online job ads to measure labor demand. Job ads and unemployment insurance claims have also been used to measure the effect of Covid-19 on the US labor market (Frosythe et al., 2020). Similarly, Hensvik et al. (2021) uses job ads postings and amount of job searches to study the effects of Covid-19 on the Swedish labor market. The study also uses data from Arbetsförmedligen, data which partly corresponds to the data used in this study. Antenucci et al. has also managed to measure labor demand using data from social media in 2014. D'Amuri and Marcucci (2009), Askitas and Zimmermann (2009) and Fondeur and Karame (2013) have successfully predicted unemployment by using online job searches. The main focus of this study is to formally evaluate if job ads can be used to measure labor demand. Many studies could benefit from this result. A good example is the paper by Hensvik et al. (2021), where the authors use digital job ads without first evaluating if these can be used to measure labor demand.

One issue in earlier literature has been that not all job seekers are equally likely to search for jobs online. A large consultancy firm might be more likely to look for labor online than a small kiosk. These methodological issues where reviewed in a study by Kureková et al. (2015). They conclude that the best way to handle them is to either statistically construct a model that compensates for the groups that might be missing from digital platforms, or compare with survey data that is randomly handed out to actors in the labor market. In this study, we use the second approach. This means that the constructed job ads index does not have to be based on all ads that exist in Sweden. It only has to cover enough jobs to be able to predict the movement of the conventional vacancy survey by Statistics Sweden.

We expect our Swedish data to be less selective than data considered in previous studies, since we have access to the centralized public employment service Arbetsförmedlingen (hereafter AF). At 70%, the percentage of job seekers who use public employment services is higher in Sweden than in most other EU countries (European commission, 2017). Swedes are also frequent users of internet services. 95% of Swedes use internet and 98% have access to it according to a recent report by the Swedish Internet Foundation (Internetstiftelsen, 2019). All jobs in the public sector must be announced on AF's web page according to §6 of the Swedish regulation of employment (SFS 1994:373) making these jobs well represented in the data we use. Thus, the data from AF provides a good representation of the Swedish labor market.

The remainder of this study is structured as follows: Section 2 gives an overview of the relevant literature and earlier research in this area. Section 3 provides information on the theoretical connection between unemployment and vacancies. Section 4 describes the data, defines our index and introduces our econometric model and theoretical framework. Section 5 gives our results and analysis. Section 6 concludes.

2 Literature review

Lenaerts et al. (2016) provide a review of the different web-based sources and methods for analyzing the labor market. They note that as most parts of the world have access to good internet, it becomes a good representation of the actual economy. The authors point to a study by Konetic (2014) which found that 80% of HR-teams use social media when hiring new employees. Google trends, social networks and the workplace recommendation website Glassdoor are all widely used to find employees, and are identified as good avenues for researching the different parts of the labor market. The authors conclude that online data has untapped potential for researchers analyzing the economy, especially the labor market.

A recent study in Italy shows that online job ads can be used to accurately infer vacancy rates (Lovaglio et al., 2019). They use the WollyBI project that gathers online vacancies from Italian job platforms to construct a job vacancy index. They then compare it with official survey statistics. This index is constructed by using so called "Web Scraping", where an algorithm finds and retrieves selected text bits, such as an ad, by searching through relevant websites. The authors then construct a labor demand index based on the ads that the algorithm has retrieved. The constructed index seems to correspond well with official statistics, but it still has some issues. The main problem comes with using web scraping to gather ads from different digital job platforms. Companies often put up the same ad on several platforms to maximize their reach. Web scraping algorithms have difficulties distinguishing whether or not an ad found on one website is actually the same as one already found on another, and so it keeps both. This double counting of ads can distort the resulting index as a measure of labor demand. They try to use machine learning-algorithms to mitigate this issue, but it is difficult to evaluate if this was successful. One important contribution of this study is that, by using the common job ads platform by AF, we remove this issue.

D'Amuri and Marcucci (2009), Askitas and Zimmermann (2009) and Fondeur and Karame (2013) have instead used online job searches to predict unemployment. They gather the amount of key word search terms in Google related to unemployment. These searches are made into an index that is compared to official survey statistics on unemployment to establish validity. The authors manage to construct search-based indices that successfully tracks unemployment for the United States, Germany and France respectively.

Two recent studies use job vacancy postings and job search data to analyze the effect of the Covid-19 pandemic on labor demand. Hensvik et al. (2021) use the amount of clicks on job ads on the Swedish job ads platform AF to see how search patterns have changed amongst people looking for jobs. They found a reduction in job search intensity right after heavy restrictions where placed on society in April 2020, measured as a decrease in clicks on job ads per user. As data and clicks are updated in real time, this study contributes with a new way to measure labor demand dynamically. In the United States, Forsythe et al. (2020) look instead at the reduction in digital vacancy postings as a result of the Covid-19 pandemic. They find that the amount of vacancy postings decreased by 40% between February and April 2020. The study also has the advantage of providing close-to real time data, but would have been benefited by a more thorough investigation of how well digital job ads can actually measure labor demand.

Another approach has been to analyze the labor market using data from social media. Antenucci et al. (2014) use a significant amount of so called "tweets" from the communication platform Twitter to predict unemployment. This is done by finding and analyzing keywords in each tweet. They find that their index seem to capture unemployment in a way similar to official statistics such as jobless claims. They also construct a modern Beveridge curve based on their findings, which resembles the theoretical predicted curve. The challenge with this kind of data is that it is not self-evident what should count as a relevant tweet or keyword. If you train a machine learning-algorithm to do this evaluation for you, it is difficult to determine if the algorithm has made this distinction correctly. The advantage of using digital job ads is that there is no ambiguity as to what information is relevant, as it is simply the stated number of vacancies.

A study using digital job ads to establish if they can be used to measure labor demand has not been conducted in Sweden. This is despite the fact that AF is providing a common platform for a large amount of national job ads, which removes the problem of duplicates present in earlier literature. This type of data is important for central bank policy makers, as labor market tightness is one crucial factor when deciding on how high the policy rate should be (Riksbanken, 2020).

3 Economic theory of vacancies and unemployment

Constructing an index that provides more information on vacancies also enables further studies of unemployment and its connection to vacancies. This can in turn be used by policy makers looking for early warning signs that unemployment will increase, and new methods for evaluating employment programs. This use of a vacancy index necessitates a relevant theoretical framework.

A common way to model the relationship between the vacancy and unemployment rate is by using the Beveridge curve, developed originally by Christopher Dow and Leslie Arthur Dicks-Mireaux (1958). The model is based on a matching model outlining how efficiently jobs and job seekers match, given labor heterogeneity. The theory posits a non-linear convex relationship where a decrease in vacancies from high levels quickly increases unemployment but at a gradually diminishing rate. A graphical representation of this theoretical relationship can be viewed in Figure 1. This figure implies that a higher vacancy rate means increased matching and so a lower rate of unemployment, in line with intuition. This effect has also been observed empirically (Olsson, 2012).

An extensive review by Elsby et al. (2015) presents frameworks of aggregate labor market analysis, with the Beveridge curve as the fundamental stepping stone. The Diamond-Mortensen-Pissarides framework models unfilled vacancies as belonging to unemployed workers engaged in a search process, and so the resulting Beveridge curve represents the matching technology that brings together job-seekers and employers. The advantage of this model is that it shows that, during a recession, if fewer vacancies are created then this leads to lower mobility on the labor market and unemployment. This has long been the standard workhorse model in the subject of modelling the labor market.

The review also highlights how empirical studies have put this model into question. For example, the persistent cyclical variation in unemployment remains a mystery, and this has started a process of reformulating the Beveridge curve in more microeconomic terms. With this study, we present another way to test the dynamics of the Beveridge

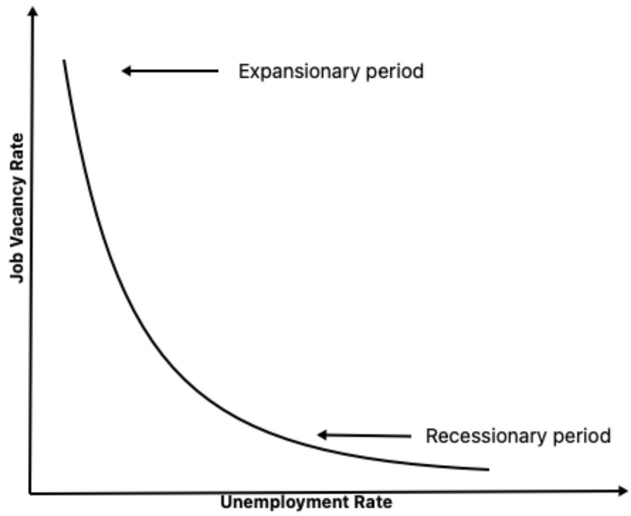

Figure 1: *A Beveridge curve showing the theoretical relationship between vacancy and unemployment rates.*

curve empirically on a microeconomic level.

4 Data & methodology

In this section we present our data sources, define our index and variables and present descriptive statistics. This is followed by the methodology used for constructing and evaluating the index.

4.1 Data sources

This study uses two main sources of data. First, all job ads come from JobTech Development (2021), an initiative by AF to make labor market data more readily available. Second, official labor force data and occupational group classification (SSYK) is provided by Statistics Sweden, the authority responsible for Sweden's official statistics.

(1) JobTech Development (i.e. AF): The job ads dataset consists of 6.7 million ads which constitutes all ads published on AF's job ads platform between the beginning of 2006 through to the end of 2020. To post job ads through AF, the employer is required to have an organizational number for identification. Each ad contains a variety of information such

as job ad headline, description, number of vacancies, employment type, scope of work, experience requirements, publication and last application date. Due to the significant amount of data, ads were stored in a relational database where analysis was conducted. For greater insight into how each job ad is structured, an example ad is presented in Table 6 in the Appendix.

(2) Statistics Sweden: To evaluate the precision of our job ads index, data is collected from Statistics Sweden's Job Openings and Unmet Labor Demand Survey (Jan. 2021). The survey collects answers from private firms, authorities and non-profit organizations. Survey participation is required by law, thus dropout rates are low. We use the survey variables vacancies and the derived measure of vacancy rate in this study. The two measures are defined as follows:

- **Vacancies:** The aggregate number of available jobs for employers at the day of measurement, where external recruitment has begun but none have yet been appointed. External recruitment refers to recruitment of individuals from outside the given organization.

- **Vacancy rate:** The relationship between the number of available jobs and the total amount employed produces the vacancy rate. This is a relative measure where available jobs have been normalized with respect to the total number employed in a group. A high vacancy rate indicates that the recruitment activity is high.

$$\text{Vacancy rate} = \frac{\text{Vacancies}}{\text{Total number employed}} * 100$$

To calculate vacancy rates for specific occupational groups, we also collect the total number of employed individuals disaggregated by occupational group from the Labor Force Survey (LFS) produced by Statistics Sweden (Mar. 2021). In contrast to the vacancy survey, LFS studies the supply side of labor demand by surveying individuals on the labor market.

To disaggregate our index by occupational group, we utilize the official occupational group classification system SSYK. This system is designed in four levels, with level one being the most aggregated level where occupations are divided into ten different categories corresponding to all occupational groups on the labor market. At level four, its most granular stage, occupational groups are divided into 429 different groups including different levels within the same occupation. For example, IT-directors level one infers the highest ranking IT-director and level two infers a subordinate position with the same title.

We will present labor demand by occupational group on SSYK level one. For researchers looking to study demand of a specific occupational group on a more granular level, it is possible to disaggregate further.

4.2 Index definition

We start our discussion of the job ads index by defining it. Below is a representation of the procedure through which the index is created.

$$i = ad,\ t = time$$

$$job\_ads\_index_t = \sum_{i=1}^{n} vacancies_i * active_{it}$$

$$active = \begin{cases} 1, & \text{if ad } i \text{ is } active \text{ in } t \\ 0, & \text{otherwise} \end{cases}$$

The sum of vacancies for all active ads is calculated for each day t from the beginning of 2006 through to the end of 2020. An ad i is *active* if t is within the range of dates between publication date (lower bound) and application deadline (upper bound) for the given ad. The attribute *active* is checked for all ads i up until n where n is the total number of ads (6.7 million). For each date in the date range, active ads are gathered and the aggregate number of vacancies summed to create the job ads index. If an ad was only active for one day, its advertised number of vacancies would only count towards the aggregate number of vacancies for that day. For the main analysis, all ads are treated equally. When disaggregating, ads are divided into specific groups where the aggregate number of vacancies is calculated based on group affiliation such as occupational group or municipality type.

4.3 Variable selection

Construction of the index begins with variable selection. As seen in the example ad in Table 6 in Appendix, each ad contains 33 variables. The number of missing values varies greatly for each variable, which is a result of two different factors. First, employers are not required to fill in all information about the job leading to some ads being more complete than others. Second, data quality has increased over the years leading to fewer missing values among some variables. The increase in quality can be attributed to advancements in AF's treatment of the job ads, where new systems automatically enrich ads with information connected to the given employer's organizational number.

Employers can post ads in two different ways. Either they can post them manually through their web browser or connect their local recruitment systems to AF's API-service and post ads directly. The possibility of direct and automatic job postings may partially explain the increase in job ads and data quality in recent years. For both the job ads index and our disaggregations, only ads that have the all the necessary information will be included.

By definition, the job ads index calculates the number of vacancies for each day between 2006 and 2020. To achieve this task, four variables are used:

- **Ad headline:** The headline of the ad. To review our ads and study various anomalies, the ad headline is used.
- **Number of vacancies:** The total number of advertised jobs per ad. The value is used to aggregate the number of vacancies among all active ads for a given day.
- **Publication date:** The date the ad was posted on AF's job ads platform. The publication date is used as a lower bound for active ads.
- **Application deadline:** The last day the ad was open for application. The application deadline is used as a upper bound for active ads.

In addition to the job ads index we also disaggregate our index to study changes in labor demand for different occupational groups and municipality types. To perform the disaggregations, two additional variables are used:

- **Municipality code:** Each municipality in Sweden has its own unique municipality code. For example, Stockholm has 0180 and Gothenburg has 1480 etc. A municipality classification system has been created by The Swedish Association of Local Authorities and Regions (2021). This system groups municipalities depending on certain properties of the municipality. For details on the specific properties of each group, see the link in the reference above. We analyze how labor demand has developed amongst the different types of municipalities according to this classification system.

- **Occupational group code:** Occupational group code on the fourth level of SSYK. The code consists of an integer value with a digit length corresponding to it's respective level i.e. the fourth level is a four digit integer value. Only the fourth level of SSYK is available in the job ads, thus we enrich this data with the complete SSYK classifications. This enables us to aggregate up to the more general levels of three, two and one, with level one containing the most broad definitions of occupational groups, which we will use in our analysis.

As municipality code and occupational group code are simply numbers, these resources in the ads are enriched with data from Statistics Sweden (complete SSYK classification) and The Swedish Association of Local Authorities and Regions (municipality types) before analysis is conducted.

4.4 Descriptive statistics

Only ads that have complete information will be included in the analysis. An ad is only considered complete if all required variables are present. For the job ads index the ad headline, number of vacancies, publication and application date need to be present for the ad to be considered complete. The disaggregations require one additional variable each. Thus, the number of complete ads will differ for the job ads index and the disaggregations by occupational group and municipality type. In Table 1 we can study the number of complete ads for each analysis.

As seen in Table 1, almost all ads contain the information needed to construct the job ads index (96.41%). For the disaggregations, fewer ads are complete. For the disaggregation by municipality type, ads are mostly complete with the exception of 2019 where a relatively greater amount is missing. In discussions with AF, no clear reason for this anomaly has been found. AF are in the process of improving the quality of the data and complete information will be available for future research. Looking at the disaggregation by occupational group, we find that almost all ads are incomplete for 2006 and 2007, as such the time frame of this analysis is limited to 2008 until 2020. It is clear from studying

	Job Ads Index			Municipality Type			Occupational Group		
	Valid	N/A	Valid(%)	Valid	N/A	Valid(%)	Valid	N/A	Valid(%)
2006-2007	479,579	49,753	90.6	468,576	60,756	88.52	34,266	495,066	6.47
2008-2009	428,767	32,580	92.94	414,581	46,766	89.86	323,300	138,047	70.08
2010-2011	612,015	41,927	93.59	598,437	55,505	91.51	522,748	131,194	79.94
2012-2013	721,501	47,910	93.77	701,597	67,814	91.19	643,564	125,847	83.64
2014-2015	1,022,201	43,297	95.94	1,003,373	62,125	94.17	981,235	84,263	92.09
2016-2017	1,407,594	25,743	98.2	1,387,609	45,728	96.81	1,403,211	30,126	97.9
2018-2019	1,323,226	0	100	1,174,154	149,072	88.73	1,321,112	2,114	99.84
2020	490,618	95	99.98	449,857	40,856	91.67	490,512	201	99.96
Total:	6,485,501	241,305	96.41	6,198,184	528,622	92.14	5,719,948	1,006,858	85.03

Table 1: *Total available ads for each analysis. The three columns for each analysis represent number of valid ads, incomplete or erroneous ads (N/A) and the share of valid ads. The dataset consists of 6,726,806 ads.*

Table 1 that data quality has increased substantially over time.

Studying the complete ads for the job ads index, we find that 78.7% of the ads only advertise one job (i.e. the variable number_of_vacancies = 1). Ads with more than one advertised job constitute a decreasing share of the total amount of complete ads. 98.8% of the job ads advertise 10 jobs or less.

Using the ≈6.5 million ads with complete information, the job ads index is created through the procedure in section 4.2. In Figure 2 we present the finished job ads index visually in daily frequency, a resampled quarterly frequency and a trend line.

Studying the index we find that there is a clear seasonal pattern. For every year except 2015, The first quarter of each year has the highest number of vacancies. In the following second and third quarter we see a decrease in vacancies followed by an increase in the fourth quarter. Looking at the trend, there is a general upward pattern in the index with a large rise around 2015-2019 which is an effect of Sweden's booming economy during this period (Konjunkturinstitutet, 2020). We also see two trend-breaking patterns in connection to downturns, specifically the 2008/2009 financial crisis and the more recent 2020 Covid-19 pandemic.

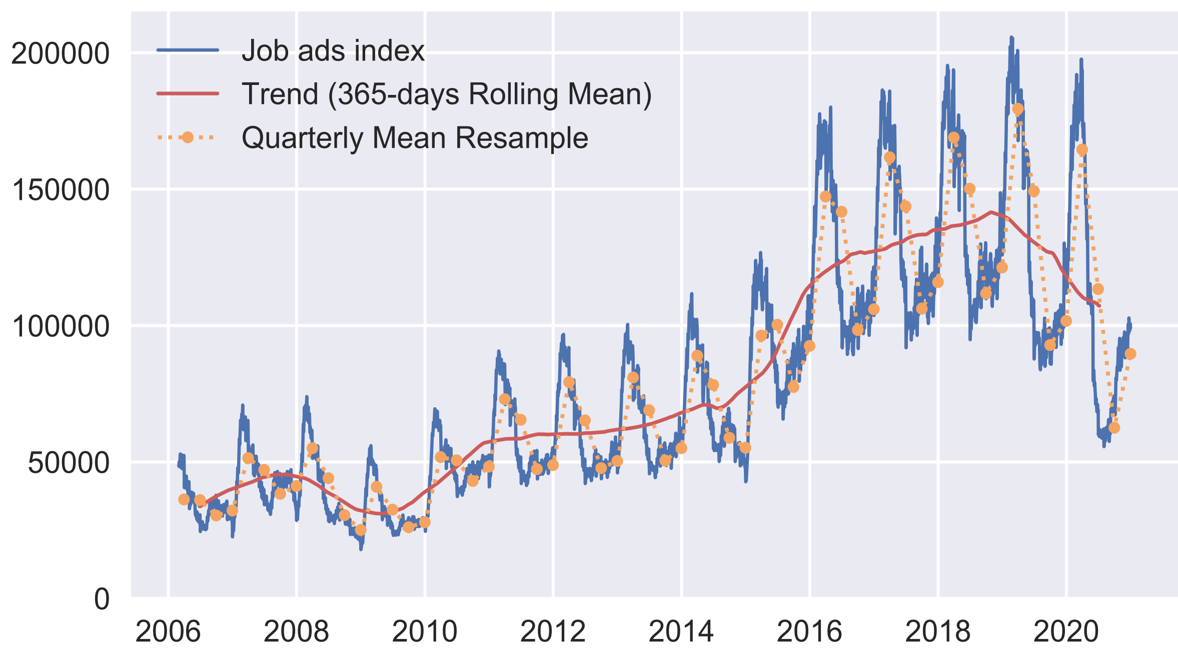

Figure 2: *The Job Ads Index. The index is presented in daily frequency, quarterly frequency and trend.*

4.5 Theory & methodology

To determine how well the job ads index corresponds to the vacancy survey, we perform a regression where the vacancy survey is the dependent variable and the job ads index is the independent variable. More precisely, we estimate the following model:

$$Vacancy\_survey_t = \alpha + \beta Job\_ads\_index_t + \epsilon$$

If this test is significant at the 1%-level, we have an indication that the two variables measure approximately the same phenomena: labor market demand. It is important to note that we do not expect these indices to be exactly the same, as they measure somewhat different things. For example, the survey asks firms how many people they have begun recruiting externally, whereas our job ads index measures a labor need already decided upon and solely based on online ads.

It should also be noted that this is not a causal analysis. Conclusions from this regression test do not imply that there is a causal relationship between our job ads index and the quarterly survey by Statistics Sweden. It simply implies that there are good grounds to assume that the two indices measure the same phenomena, which is some sort of labor demand. Therefore, we will mainly focus on how well the model explains the variation in the vacancy survey.

Since the two indices do not share the same frequency, they cannot be estimated using OLS in their original form. Therefore, we resample the daily job ads index into quarterly frequency by taking the average value over each quarterly period and assign this to each quarter. This quarterly version of the job ads index will only be used in this OLS setting. The quarterly mean resample of the job ads index is represented visually in Figure 2. Similarly, when creating the Beveridge curves, the job ads index will be compared to monthly unemployment rates, and so for this purpose the job ads index will be resampled to monthly frequency using the same average-over-period method.

Table 2 provides summary statistics for the two main variables used in the regression analysis. The min and max values do not indicate that there are any outliers regarding the number of vacancies each quarter. The mean value is similar between the two indices, but the standard deviation is larger for the job ads index suggesting that it captures more of the volatility in the labor market.

Since these two indices are time series, an issue that could give a spurious relation is non-stationarity, either due to a trend-component affecting their growth, or if there exists a unit-root process in any of the time series. Since the number of vacant positions, unlike for example stocks, are naturally bounded by the number of firms and individuals in an economy it is not likely that there is a unit-root process in the data. If there is a time trend it justifies performing a first-difference to detrend the data and make it stationary.

To find if either a unit-root or an underlying trend component affects our data we perform a Dickey-Fuller test of non-stationarity (Dickey & Fuller, 1979). This test estimates the following model:

$$\Delta y_t = \alpha + \delta y_{t-1} + \beta t + \epsilon$$

Where y_t is either the vacancy survey or the job ads index, Δ is the first difference and t is a deterministic time trend, while ϵ is random noise and α is a constant controlling for drift. $\delta = \rho - 1$ is a coefficient we test if different from zero, since if $\delta = 0$ then y_{t-1} is not able to predict y_t. This means there is a random element in the time series' growth over time, indicating a unit-root. β is a coefficient which controls for the fact that the relation might be governed by a long-term growth trend. This test will be performed with and without the time trend, to determine if the trend component is the likely cause of non-stationarity. Should the regression without the time trend be significant at the 1%

level we can reject the null hypothesis that the process is non-stationary, and estimate the main regression model using unmodified versions of the vacancy survey and job ads index. If it is significant only when the time trend is included in the model, then a first difference is necessary to remove this trend. If neither of the two models are significant then further differences such as the second difference need to be performed. In the final model we also include an interaction term between a dummy variable for each quarter and the job ads index. This is to determine if the index captures variation in labor demand only during certain periods, or if its performance is equal throughout each year.

We use LOESS to examine the trend and seasonal fluctuations of both indices. LOESS is short for locally estimated scatter plot smoothing and is a non-parametric regression technique that fits a line based on a weighted moving time window. This will decompose a time series index into a trend, seasonal and residual component (Cleveland, 1990). This is interesting since in the social media study by Antenucci et al. (2014) the seasonal component was not similar between the constructed and the official index. In the Italian study by Lovaglio et al. (2019) the authors did find a similar seasonal component in both indices. It can therefore be interesting to compare trend and seasonal variation between the job ads index and Statistics Sweden's vacancy survey. If we find similar trend and seasonal components in the job ads index and the vacancy survey, it would further support that they are affected by the same underlying forces.

After the main analysis is complete we disaggregate the index by municipality type and occupational group. This enables us to plot labor demand over time for different municipality types and occupations. These disaggregations are meant to show possible use cases of the job ads index, and how it can contribute to further studies of labor demand.

To establish a connection between the job ads index and unemployment rate we construct a Beveridge curve out of the data. The curve plots the unemployment rate against the vacancy rate. According to classical theory outlined under section 3, the result is expected to be a form similar to the convex upward line displayed in Figure 1, where lower rates of unemployment correspond to a higher vacancy rate, i.e. higher labor demand. This is followed by occupation-specific Beveridge curves where unemployment is plotted against vacancy rates for SSYK level one aggregated occupations.

Variables	(1) N	(2) mean	(3) sd	(4) min	(5) max
Job ads index	60	78,271	41,978	25,097	179,465
Vacancy survey	60	74,830	29,829	27,605	143,769

Table 2: *Summary statistics for the two time series, Statistics Sweden's vacancy survey and the job ads index, included in the main regression and the Dickey-Fuller tests. Both series are in quarterly frequency, meaning the job ads index was resampled.*

5 Results & analysis

In this section, we examine possible non-stationarity and regress Statistics Sweden's vacancy survey on the job ads index. We then present the results of our disaggretations and the constructed Beveridge curves based on the digital job ads index. Each analysis is aided by a graphical representation.

5.1 Comparing the job ads index with the vacancy survey

As a first step in evaluating the job ads index we examine it visually together with Statistics Sweden's quarterly survey. In Figure 3 we see that the two time series show a clear co-movement. This indicates that much of the variation in the official survey might be explained by the daily index.

Before regressing the vacancy survey on the job ads index, we need to establish if any of the two indices are non-stationary by performing a Dickey-Fuller test. The result of this test is shown in Table 3. Beginning with the vacancy survey, the initial DF-test is insignificant, indicating non-stationary. When controlling for a time-trend, it becomes significant at the 1%-level, suggesting that the non-stationarity is due to a trend. Performing the DF-test on the first-difference of the vacancy survey results in the index being significant at the 1%-level, and so the first-difference in the vacancy survey should be used for the main regression model.

A DF-test of the job ads index reveals a similar result. This index is significant at the 10%-level without the time trend included, and significant at the 1%-level when a time trend is controlled for. When performing the DF-test on the first-difference of the job ads index instead both models become significant at the 1%-level. Therefore, the first-difference version of the job ads index will also be included in the main regression

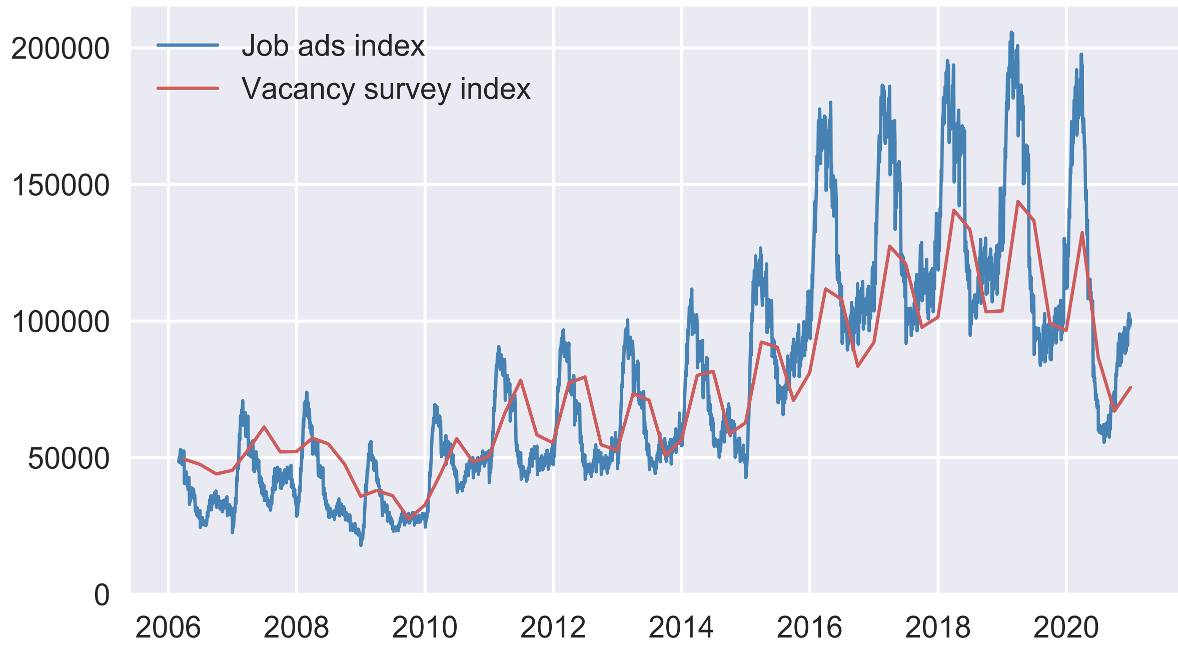

Figure 3: *Aggregate number of vacancies between 2006-2020 for the job ad index and the vacancy survey by Statistics Sweden.*

Variables	(1) no trend	(2) trend
Vacancy survey	0.1221	0.0080***
Survey diff	0.0000***	0.0000***
Job ads index	0.0839*	0.0009***
Ads diff	0.0000***	0.0000***

MacKinnon approximate P-values
*** p<0.01, ** p<0.05, * p<0.1

Table 3: *Dickey-Fuller tests of non-stationarity. The models are each variable regressed on themselves after applying first-difference, with or without a deterministic time trend. The number of observations are 60 without first-difference applied, and 59 when applied.*

model.

On the left side of Figure 4 we present the indices with the first-difference applied. It is clear that any trend component has been removed by the operation. On the right we see a scatter plot where each observation represents the relative change of the two indices. The distance of each observation from the fitted line tells us to which degree the change from one period to the next coincides between the indices. Since the overall residuals are low, the correlation between the two indices is high.

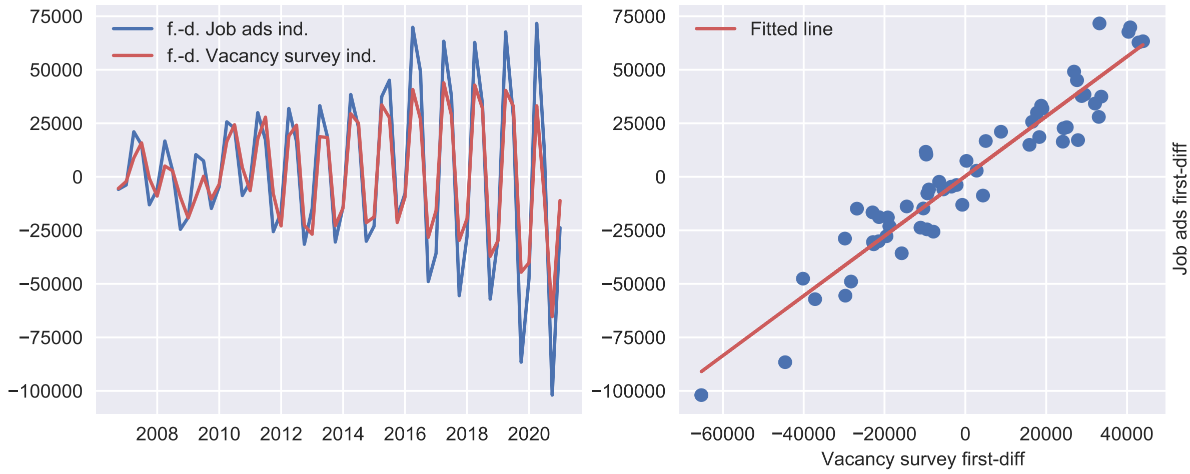

Figure 4: *The first-difference of the job ads index and vacancy survey as a line plot over time (left) along with a scatter plot showing the relative change of the two indices with a fitted line (right).*

We regress the vacancy survey on the job ads index re-sampled to quarterly frequency. The result is displayed in Table 4. Model 1 is the result from the regression using the original form of the two indices without first-difference applied, and here the job ads index is significant at the 1% level, with a very high R^2 of 0.96. Model 2 is the same regression but with first-difference applied to the two indices for trend removal. The job ads index remains significant at the same level. The R^2 is now slightly lower at 0.87, although it can still be considered high.

In the third model the first-difference of the job ads index is interacted with each quarter, comparing quarter two, three and four with quarter one. This is to determine if the job ads index explanatory power is dependent on which quarter we are considering. This does not seem to be the case for the second and third quarter, which are both insignificant. The fourth quarter is significantly different from the first at the 5%-level. Presumably, this is because the job ads index captures more seasonal variation than the vacancy survey, and the seasonal difference is the most stark between the end and beginning of each year. This is also visible in Figure 3 and will be discussed below.

The estimated coefficient in all models is around 0.6, meaning that for every new vacancy in the job ads index only 0.6 to 0.7 vacancies are added to the quarterly survey. This confirms that the job ads index generally captures more vacancies than the quarterly vacancy survey, with the caveat that these estimations might be sensitive to the method

Variables	(1) Survey vacancies	(2) Survey diff	(3) Survey diff
Ads diff		0.626***	0.624***
		(0.0319)	(0.0423)
2.quarter#c.Ads diff			-0.0685
			(0.253)
3.quarter#c.Ads diff			0.0588
			(0.111)
4.quarter#c.Ads diff			-0.251**
			(0.123)
Job ads index	0.695***		
	(0.0200)		
Observations	60	59	59
R-squared	0.956	0.867	0.874

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: *Regression models with Statistics Sweden's vacancy survey regressed on the job ads index, and the first-difference of both series labeled Ads diff and Survey diff respectively. The variables quarter 2 to 4 are dummy variables for each quarter. All series are in quarterly frequency, meaning the job ads index was resampled.*

for re-calculating to quarterly frequency. Most importantly, the fact that the job ads index remains significant at the 1% level with a high R^2 in all models indicates that this index seems to measure the same phenomena as the vacancy survey. Thus, the job ads index can safely be used as a complement to measure labor demand.

A chart comparing the seasonal and trend components for both indices can be viewed in Figure 5. Both indices have very similar trends and seasonal components until after 2014 where the job ads index starts showing larger trend- and seasonal patterns. There could be several explanations for this. As internet usage increases and more job ads are posted online the dynamics of job searching changes. The decline of traditional recruitment channels such as newspapers and local employment offices is an illustrative example. In 2019, AF closed down 132 local employment offices which further emphasizes the shift towards an increasing focus on online job searching (SVT, 2019). This systemic change could lead recruiters to move their recruitment processes online, which would make online job ads better reflect the seasonal volatility that exists among jobs in Sweden.

It should also be noted how the seasonal fluctuations in the job ads index generally appear somewhat earlier than the vacancy survey. This suggest that the job ads index by

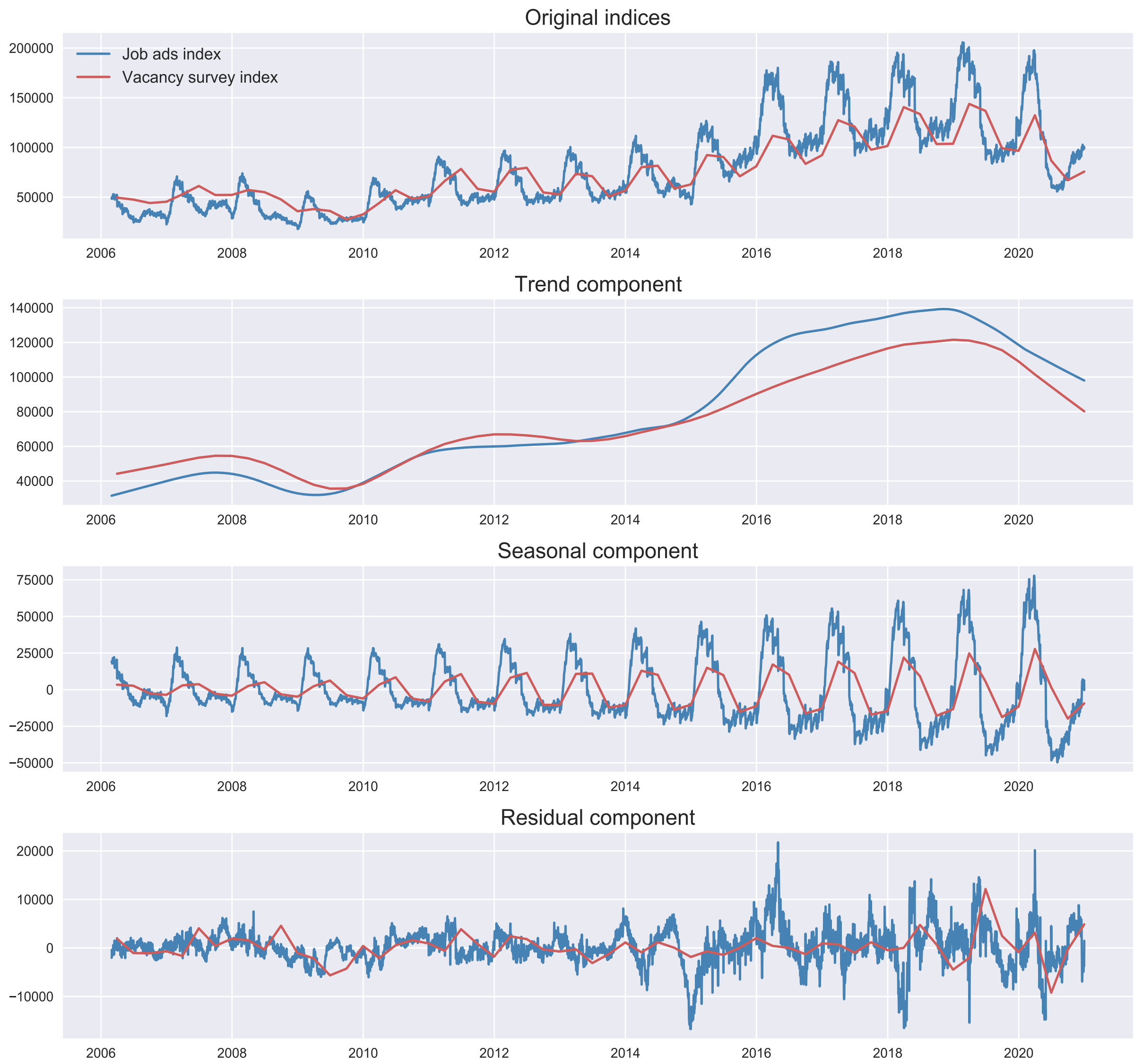

Figure 5: *Seasonal, trend and residual decomposition using LOESS, plotted together with the original version of the indices for reference.*

nature of being of a higher frequency captures changes in labor demand earlier than the vacancy survey, which instead aggregates spikes and through over a three-month period. A similar explanation can be given to the difference shown in their residual component also featured in Figure 5.

The high frequency of the job ads index captures large recessions that quickly recovers in the same period, which in the vacancy survey becomes a more modest dip in labor demand when captured in quarterly frequency. This comparison of each individual component reinforces the theory that the two indices measures the same phenomena. The main difference is that the job ads index captures changes in labor demand that are lost when aggregating over a quarter.

5.2 Disaggregations

In Figure 6 we find the result of the disaggregation on municipality type. Overall, there is a clear seasonal pattern for all municipality types which reflects the pattern we observed in the job ads index. Similarly, we see an upward trend for most municipality types although not all, and of a varying positive gradient. The largest increase in vacancies has been in large cities. This can be due to higher growth and so higher demand for labor, but also that firms located in larger cities are more likely to use digital job ads to look for labor. As a comparison, the change has been negligible in rural municipalities that have a visitor industry. This could either be due to low growth in these kinds of municipalities, or that industry centred around tourism in rural municipalities tend to recruit labor among friends and relatives, and not using digital job ads.

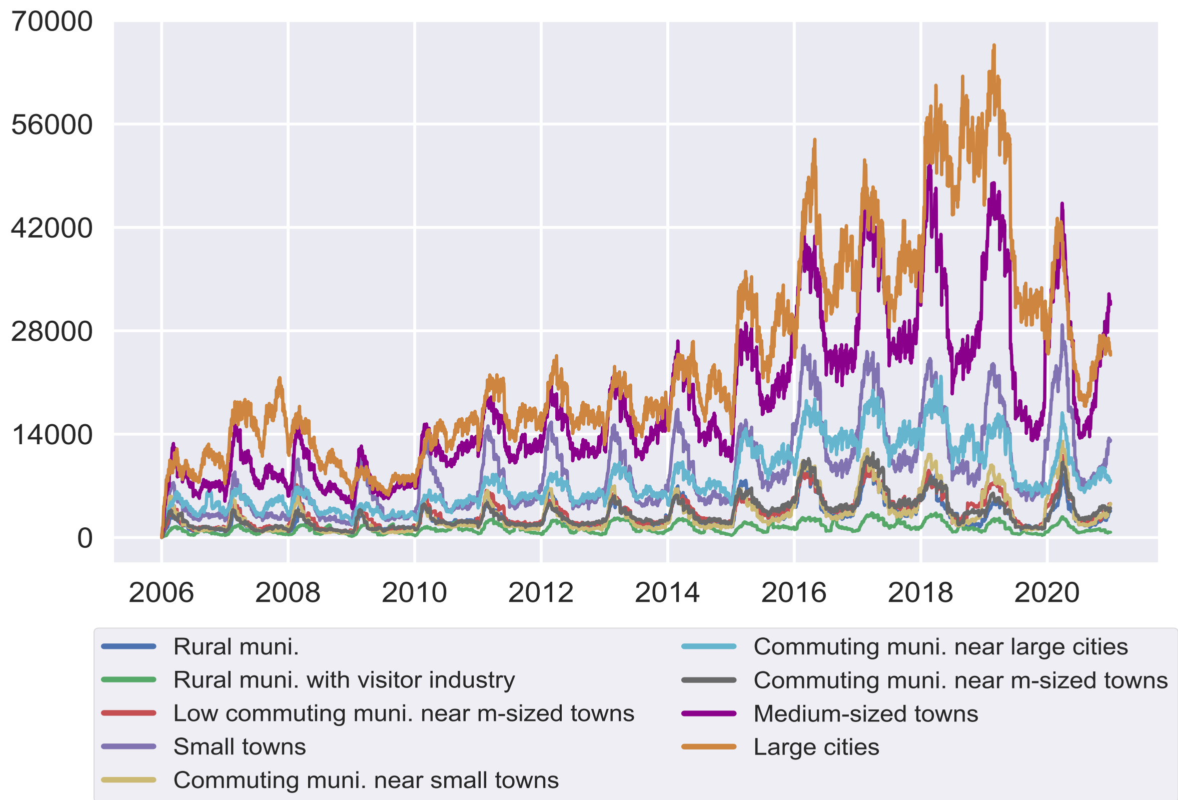

Figure 6: *Labor demand by municipality type 2006-2020. 'muni.' and 'm-sized' are abbreviations of municipality and medium-sized respectively. For properties of each municipality type see Swedish Association of Local Authorities and Regions (2017).*

The disaggregation by occupational group on SSYK level one, the most aggregated level, can be examined in Figure 7. Note that due to lack of observations the group military occupations has been excluded. The most noticeable group is the demand for service workers. This type of labor is heavily influenced by seasonal fluctuations, and contains two of the most common occupations in Sweden (Statistics Sweden, 2020) which are assistant nurses and shop assistants. It can therefore be assumed that as more public sector organisations and private firms use digital job ads to look for labor, our constructed index captures these strong seasonal fluctuations with increasing accuracy. In comparison, firms under the heading of agriculture, horticulture, forestry and fishery have barely increased at all since 2006. This could be explained by the fact that firms and organisations in this sector do not use digital ads to look for employees. They might prefer to use contacts such as people they know in their town, and may to a higher degree than other sectors contract labor from outside Sweden, which will not visible in digital job ads on AF.

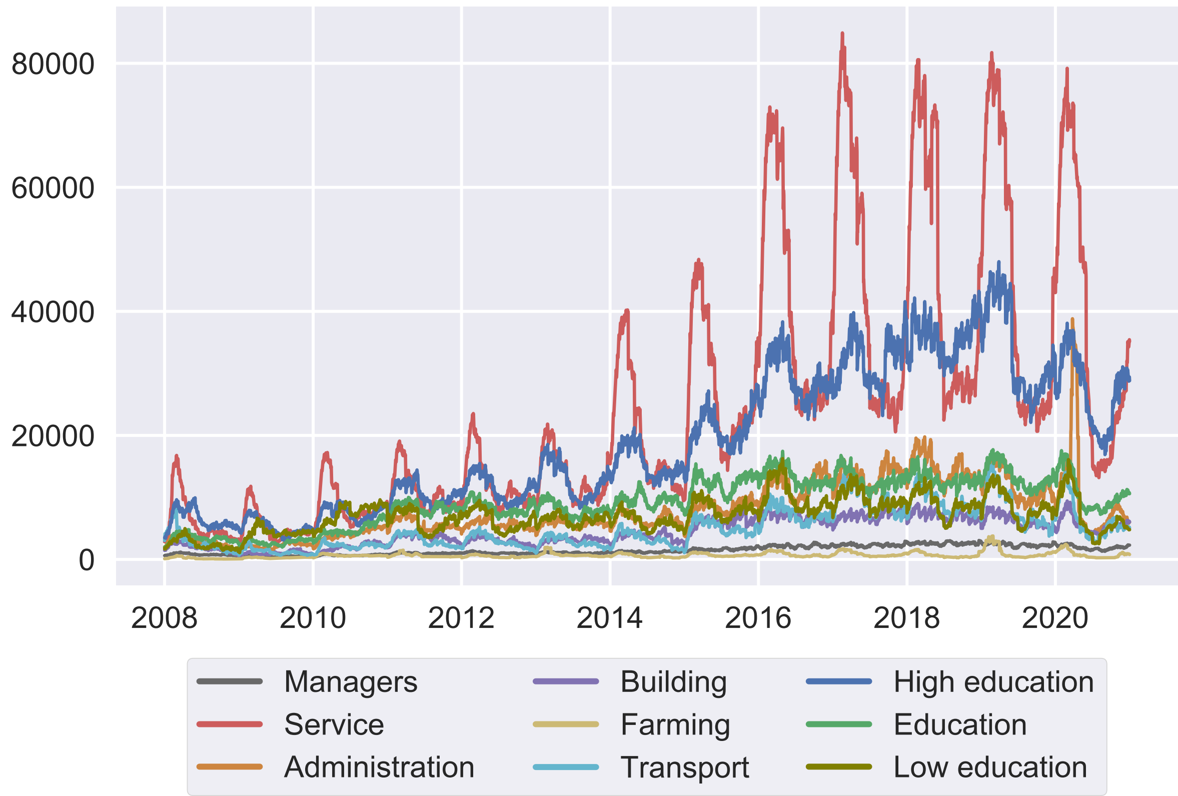

Figure 7: *Labor demand by occupational group on SSYK level one, 2008-2020. Demand for military personnel has been excluded due to lack of observations. For full definitions of each group see SCB (2017).*

5.3 Beveridge curves

In comparison to the vacancy survey, the job ads index enables examination of the relationship between vacancy rates and unemployment rates both using more observations, and specifically among different occupational groups. In Figure 8 we have plotted Statistics Sweden's vacancy survey against quarterly unemployment rate for the period between 2010 to 2020. This shows a clear, negative linear relationship between the two variables similar to what was expected from the theoretical Beveridge curve in Figure 1, although not convex. This chart however has only a limited number of observations. Data prior to 2010 was omitted to avoid the great financial crisis of 2008 acting as an outlier.

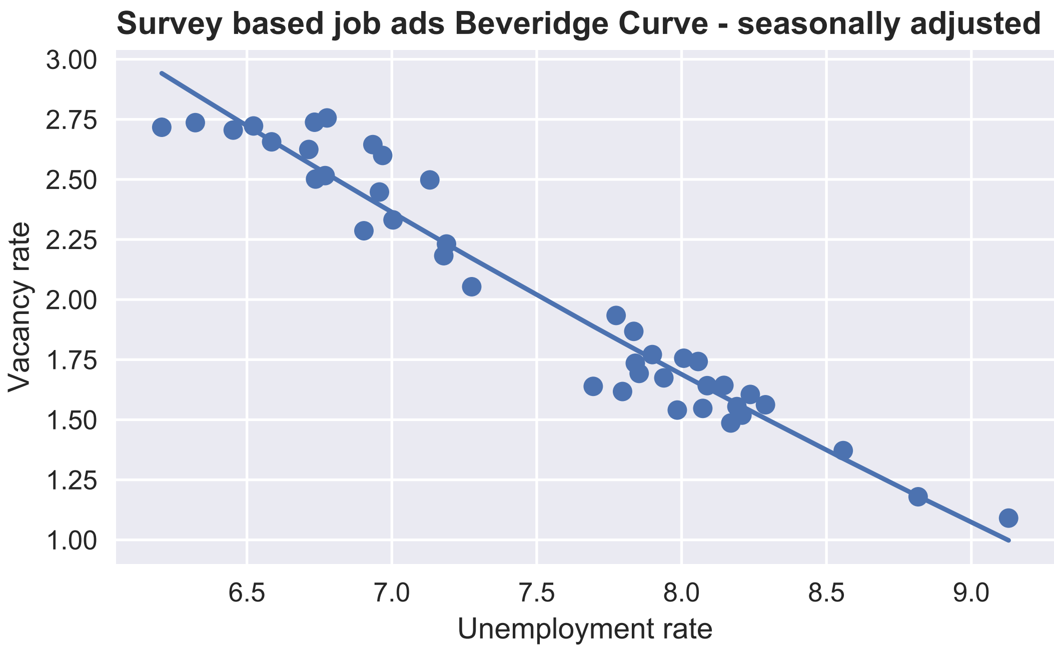

Figure 8: *Seasonally adjusted Beveridge curve based on Statistic Sweden's quarterly vacancy survey and quarterly unemployment rate, with a fitted line. The data covers the time period 2010 Q1 to 2020 Q1.*

Figure 9 plots the same relationship but with the job ads index, scaled to monthly frequency against monthly unemployment rate for the same period. This chart also shows a negative relationship but with a slight concave line. Both charts suggest, in line with theory, that a higher vacancy rate is related to low unemployment as there are more positions for unused labor to apply for. As the vacancy rate decreases, the unemployment rate increases. The difference between this empirical relationship and the theoretical one is that it is linear: a decrease in the vacancy rate is not followed by first a rapid then diminishing increase in unemployment, and vice versa. Both these charts have been seasonally adjusted and the non-seasonally adjusted versions, showing a similar although more noisy relationship, can be viewed in the Appendix in Figure 11 and Figure 12.

In Figure 10 we have disaggregated the job ads index to occupational group level

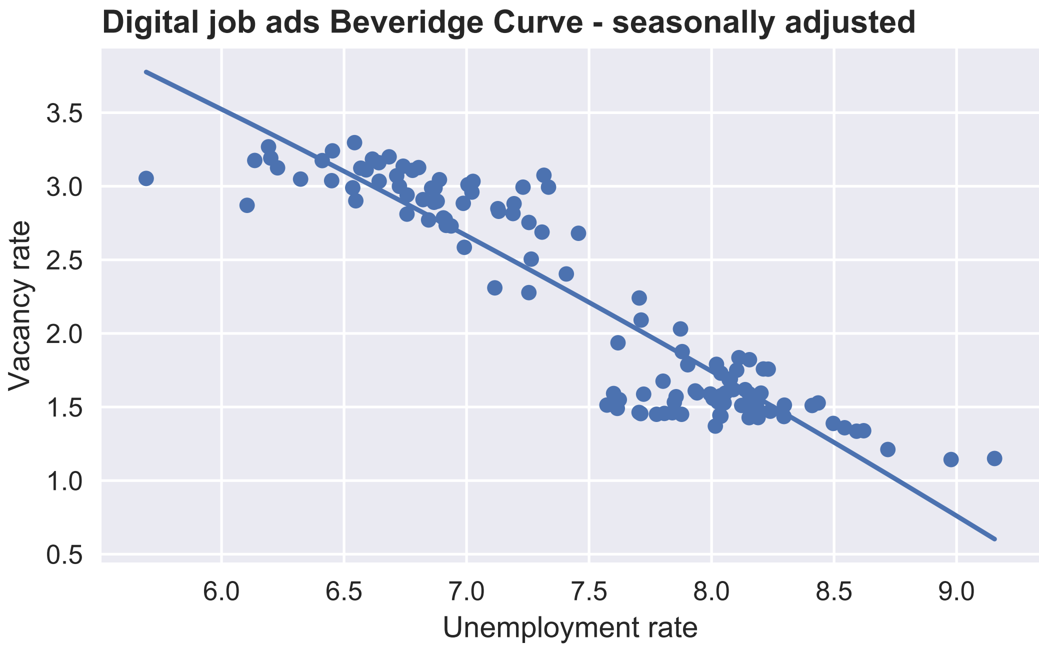

Figure 9: *Seasonally adjusted Beveridge curve based on the job ads index resampled to monthly frequency and monthly unemployment rate, with a fitted line. The data covers the time period 2010-03 to 2020-03.*

one, calculated the vacancy rate and plotted against monthly unemployment rate. This enables examination of how the dynamics between the vacancy rate and unemployment rate changes between different occupations. Due to data restrictions, this analysis is only available for the period 2015 through to 2020. The full name of each group can be viewed in Table 5 in the Appendix.

The difference between some occupational groups is stark. Work related to manufacturing and managerial work follow the same negative linear trend as the main aggregated index, whereas occupations related to administration, transportation, service and work requiring higher education show a relationship more similar to the theoretical, non-linear Beveridge curve. Work which demands elementary or some level of higher education follow a concave, non-linear relationship while work related to agriculture has an almost positive linear relationship, contrary to theoretical predictions. For work that demands higher education it should be kept in mind that these are occupations grouped at the most aggregated level, and so many different kinds of work with different dynamics are included in these measures. This is probably an explanation for the diverging observations under Service-jobs as well. Data on occupational employment is presently not available at a more granular level.

Regarding work related to agriculture, one possible reason behind the counter-intuitive relation is that firms acting in this sector could to a larger degree hire people based on family ties or friends. Digital job ads would thus not be representative of this sector's labor demand. There is also one more occupational group, military personnel, which was not included in this analysis due to a lack of observations. This disaggregation shows that it is possible to study the dynamics of unemployment and vacancy rates in more detail using a labor demand index based on digital job ads. Future research can utilize the job ads index to delve deeper into the dynamics of the Beveridge curve theory.

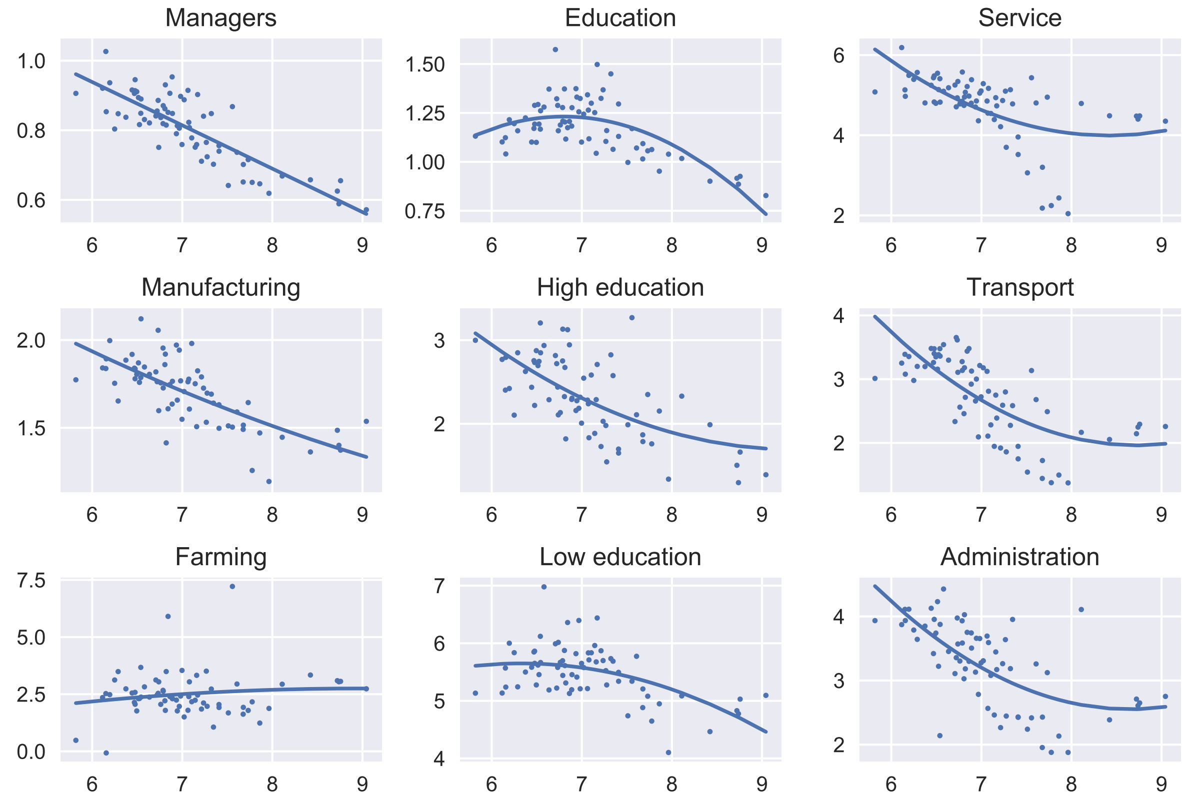

Figure 10: *Occupational Beveridge curves based on the job ads index disaggregated to SSYK level one and resampled to monthly frequency, plotted against monthly unemployment rate with a fitted line. The data is seasonally adjusted and covers the time period 2015-01 to 2020-09 due to occupational SSYK data only being available from 2015 and onward. The full name of each occupational group can be viewed in Table 5 in the Appendix.*

6 Discussion & conclusion

In this study, we have constructed an alternative measure of labor demand based on digital job ads. This index was compared to a conventional vacancy survey. We found that the job ads index tracks the vacancy survey well. This means that an index based on digital job ads can be used as a complement to official survey statistics on vacancies. The job ads index is of a higher frequency than the quarterly survey and can be disaggregated to a more detailed level, down to demand in specific regions and for specific occupations. The high frequency of the data enables policy makers to react faster to large changes in labor demand, possibly due to a recession. The increased level of detail enables researchers to study labor demand in more subtle ways.

The index is not perfect as there are issues with which ads are represented online, and how much information recruiters provide when posting their job ads. These issues are likely to be resolved over time, as AF is working actively to increase the quality of ads posted on their job ads platform. The trend of recruiting more workers online shows no sign of tapering off. This digital job ads index will therefore increase in reliability and quality as time progresses. Due to data only being available from 2006 this study focuses on a limited business cycle. Further studies should look at the interaction between unemployment and vacancies using data covering longer periods of growth and recessions.

We find that the seasonal patterns in the job ads index correspond well to the patterns in the official vacancy survey, which is similar to what Lovaglio et al. (2019) found in Italy. This stands in contrast to the study by Antenucci et al. (2014) as they found that the seasonal patterns in their unemployment index based on internet data differed from the ones found in official jobless claims. The aggregated Beveridge curve constructed from our index corresponds to the theoretical shape of the Beveridge curve. This is in line with the findings by Antenucci et al. (2014) when they used their constructed unemployment index to plot the same curve against survey vacancy data.

We end the study by constructing occupation-specific Beveridge curves modelling the relationship between unemployment and vacancies. We found that this can be an interesting avenue for further research in the subject of labor economics.

This study contributes to a growing scientific literature of using internet data to study the economy, and provides an example of how this can be done successfully using new kinds of data and methods.

7 References

Antenucci, D., Cafarella, M., Levenstein, M., Ré, C., & Shapiro, M. D. (2014). Using Social Media to Measure Labor Market Flows. (20010). https://ideas.repec.org/p/nbr/nberwo/20010.html

Askitas, N., & Zimmermann, K. (2009). Google econometrics and unemployment forecasting. (4201). https://EconPapers.repec.org/RePEc:iza:izadps:dp4201

Boselli, R., Cesarini, M., Mercorio, F., & Mezzanzanica, M. (2018). Classifying online job advertisements through machine learning. *Future Generation Computer Systems*, *86*, 319–328. https://doi.org/https://doi.org/10.1016/j.future.2018.03.035

Carnevale, A. P., Jayasundera, T., & Repnikov, D. (2014). Understanding online job ads data. *Georgetown University, Center on Education and the Workforce, Technical Report (April)*.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). Stl: A seasonal-trend decomposition procedure based on loess (with discussion). *Journal of Official Statistics*, *6*, 3–73.

D'Amuri, F., & Marcucci, J. (2009). 'Google it!' Forecasting the US unemployment rate with a Google job search index. (2009-32). https://ideas.repec.org/p/ese/iserwp/2009-32.html

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366a), 427–431. https://doi.org/10.1080/01621459.1979.10482531

DOW, J. C. R., & DICKS-MIREAUX, L. A. (1958). THE EXCESS DEMAND FOR LABOUR A STUDY OF CONDITIONS IN GREAT BRITAIN, 1946–56 1. *Oxford Economic Papers*, *10*(1), 1–33. https://doi.org/10.1093/oxfordjournals.oep.a040791

Elsby, M. W., Michaels, R., & Ratner, D. (2015). The beveridge curve: A survey. *Journal of Economic Literature*, *53*(3), 571–630.

European commission, european semester thematic factsheet [Accessed: 2020-11-25]. (2017). https://wayback.archive-it.org/12090/20201012083437/https://ec.europa.eu/info/sites/info/files/european-semester_thematic-factsheet_public-employment-services_en_0.pdf

Fondeur, Y., & Karamé, F. (2013). Can google data help predict french youth unemployment? *Economic Modelling*, *30*(100), 117–125. https://EconPapers.repec.org/RePEc:eee:ecmode:v:30:y:2013:i:c:p:117-125

Forsythe, E., Kahn, L. B., Lange, F., & Wiczer, D. (2020). Labor demand in the time of covid-19: Evidence from vacancy postings and ui claims. *Journal of Public Economics*, *189*, 104238. https://doi.org/https://doi.org/10.1016/j.jpubeco.2020.104238

Hensvik, L., Le Barbanchon, T., & Rathelot, R. (2021). Job search during the covid-19 crisis. *Journal of Public Economics*, *194*, 104349. https://doi.org/https://doi.org/10.1016/j.jpubeco.2020.104349

Internetstiftelsen, meningsfull tid på nätet – och det digitala samhällets fram- och baksidor [Accessed: 2020-11-19]. (2019). https://svenskarnaochinternet.se/rapporter/svenskarna-och-internet-2019/sammanfattning/

Jobtech development, historical jobs [Accessed: 2021-01-10]. (2021). https://jobtechdev.se/docs/apis/historical/

Konetic. (2014). *80% of hr teams use social media for sourcing and hiring new staff* [Accessed: 2020-11-19]. https://www.executivegrapevine.com/content/article/2014-09-16-80-of-hr-teams-use-social-media-for-sourcing-and-hiring-new-staff

Lenaerts, K., Beblavý, M., & Fabo, B. (2016). Prospects for utilisation of non-vacancy internet data in labour market analysis—an overview. *IZA Journal of Labor Economics*, *5*. https://doi.org/10.1186/s40172-016-0042-z

Lovaglio, P. G., Mezzanzanica, M., & Colombo, E. (2020). Comparing time series characteristics of official and web job vacancy data. *Quality & Quantity: International Journal of Methodology*, *54*(1), 85–98. https://doi.org/10.1007/s11135-019-00940-

Olsson, O. (2012). *Essentials of advanced macroeconomic theory*. Routledge. https://books.google.se/books?id=-Y0GtwAACAAJ

Riksbanken, inflationsmålet [Accessed: 2020-11-16]. (2020). https://www.riksbank.se/sv/penningpolitik/inflationsmalet/

SFS. (1994). *Anställningsförordning (1994:373)* [
http://rkrattsbaser.gov.se/sfst?bet=1994:373].

Statistics sweden, job openings and unmet labour demand [Accessed: 2021-02-14]. (2021). https://www.scb.se/en/finding-statistics/statistics-by-subject-area/labour-market/vacancies-and-unemployment/job-openings-and-unmet-labour-demand/

Statistics sweden, jobs in sweden [Accessed: 2021-02-14]. (2020). https://www.scb.se/hitta-statistik/sverige-i-siffror/utbildning-jobb-och-pengar/yrken-i-sverige/%5C#:~:text=Undersk%5C%C3%5C%B6terska%5C%20inom%5C%20hemtj%5C%C3%5C%A4nsten%5C%2C%5C%20hemsjukv%5C%C3%5C%A5rden%5C%20eller,butikss%5C%C3%5C%A4ljare%5C%20i%5C%20fackhandeln%5C%20och%5C%20grundskoll%5C%C3%5C%A4rare

Statistics sweden, konjunkturstatistik över vakanser [Accessed: 2020-11-13]. (2020). https://www.scb.se/hitta-statistik/statistik-efter-amne/arbetsmarknad/vakanser-och-arbetsloshet/konjunkturstatistik-over-vakanser-kv/

Statistics sweden, labour force surveys [Accessed: 2021-03-12]. (2021). https://www.scb.se/en/finding-statistics/statistics-by-subject-area/labour-market/labour-force-surveys/labour-force-surveys-lfs/

Statistics sweden, standard för svensk yrkesklassificering (ssyk) [Accessed: 2020-11-23]. (2017). https://www.scb.se/dokumentation/klassifikationer-och-standarder/standard-for-svensk-yrkesklassificering-ssyk/

Svt, arbetsförmedlingens besked: Här är kontoren som försvinner [Accessed: 2021-02-23]. (2019). https://www.svt.se/nyheter/inrikes/arbetsformedlingens-besked-om-kontoren

Swedish association of local authorities and regions, classification of swedish municipalities 2017 [Accessed: 2021-04-13]. (2017). https://skr.se/skr/tjanster/kommunerochregioner/faktakommunerochregioner/kommungruppsindelning.2051.html

8 Appendix

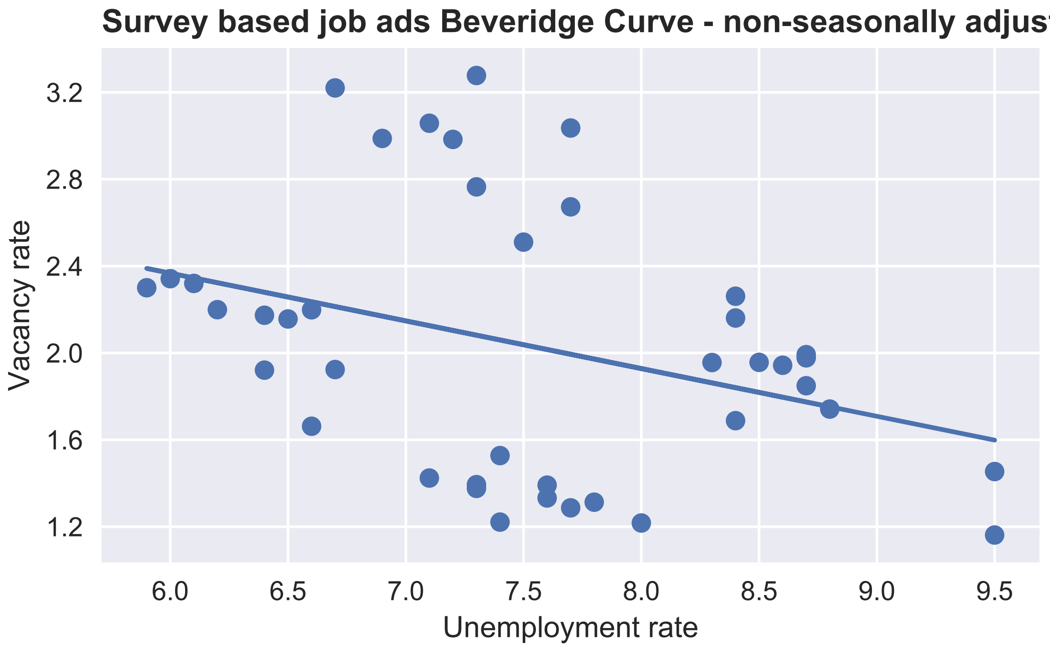

Figure 11: *Non-seasonally adjusted Beveridge curve based on Statistic Sweden's quarterly vacancy survey and quarterly unemployment rate, with a fitted line. The data covers the time period 2010 Q1 to 2020 Q1.*

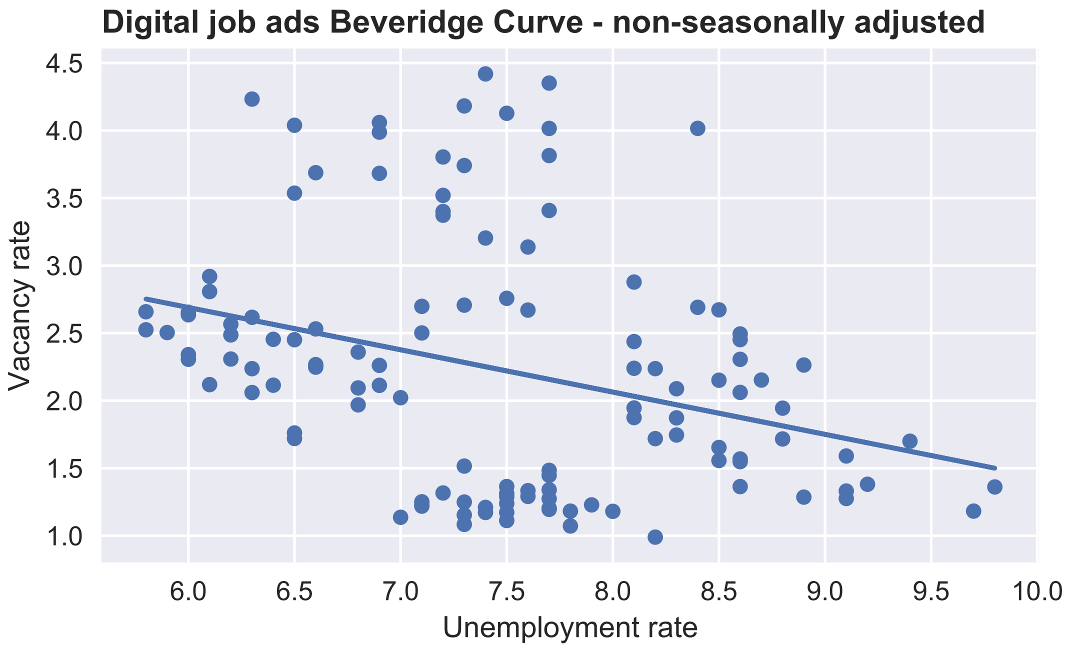

Figure 12: *Non-seasonally adjusted Beveridge curve based on job ads index resampled to monthly frequency and monthly unemployment rate, with a fitted line. The data covers the time period 2010-03 to 2020-03.*

Short name	SSYK full occupational group name
Managers	Managers
High education	Occupations requiring advanced level of higher education
Education	Occupations requiring higher education qualifications or equivalent
Administration	Administration and customer service clerks
Service	Service, care and shop sales workers
Farming	Agricultural, horticultural, forestry and fishery workers
Manufacturing	Building and manufacturing workers
Transport	Mechanical manufacturing and transport workers, etc.
Low education	Elementary occupations

Table 5: *Table for showing the long name for each occupational group under SSYK level one, which is used to plot occupational Beveridge curves in Figure 10.*

Variables	Data
access	None
access_to_own_car	None
application_deadline	2020-02-29 23:59:59,000000000
application_details	{'email: None', 'information: None', 'other': None, 'references': None, 'url': 'https://recruit.visma.com/Public/Apply.aspx?guidAssignment=f2...', 'via_af': None}
description	{'company_information': ' ', 'conditions': 'Timavlönad. Tidsbegränsad anställning, tillträde: Tillträde enligt överenskommelse. Månadsvikariat eller timanställning. Arbetstid enligt schema. Varierar mellan dag-, kvälls-, natt- och helgtjänstgöring.', 'needs': None, 'requirements': None, 'text': 'I Malmö stad arbetar vi med respekt, engagemang och kreativitet för att utveckla Malmö. Vi har Sveriges viktigaste jobb. Hos Malmö stad finns framtidens arbetsplats..., 'text_formatted': None}
driving_license	None
driving_license_required	None
duration	{'concept_id': None, 'label': 'TRE_TILL_SEX_MANADER', 'legacy_ams_taxonomy_id': None}
employer	{'email': None, 'name': 'Malmö kommun', 'organization_number': None, 'phone_number': None, 'url': None, 'workplace': 'Malmö stad Funktionsstödsförvaltningen, 177 Avd stöd, hälsa och DV'}
employment_type	{'concept_id': None, 'label': 'VANLIG_ANSTALLNING', 'legacy_ams_taxonomy_id': None}
experience_required	None
external_id	46-212000-1124-F227924B5E2F41BCBA5872B6E94D9713
headline	Välkommen att jobba som personlig assistent!
id	23725351
last_publication_date	2020-02-29 23:59:59,000000000
logo_url	None
must_have	{'languages': None, 'skills': None, 'work_experiences': None}
nice_to_have	{'languages': None, 'skills': None, 'work_experiences': None}
number_of_vacancies	5
occupation	{'concept_id': 'eU1q_zvL_9Rf', 'label': 'Personlig assistent', 'legacy_ams_taxonomy_id': '5798'}
occupation_field	{'concept_id': 'GazW_2TU_kJw', 'label': 'Socialt arbete', 'legacy_ams_taxonomy_id': '16'}
occupation_group	{'concept_id': 'sq3e_WVv_Fjd', 'label': 'Personliga assistenter', 'legacy_ams_taxonomy_id': '5343'}
publication_date	2020-01-01 00:00:00,000000000
removed	None
removed_date	None
salary_description	Enligt avtal.
salary_type	{'concept_id': None, 'label': 'FAST_LON', 'legacy_ams_taxonomy_id': None}
scope_of_work	{'max': None, 'min': None}
source_type	None
timestamp	None
webpage_url	None
working_hours_type	{'concept_id': None, 'label': 'DELTID', 'legacy_ams_taxonomy_id': None}
workplace_adress	{'city': 'Malmö', 'coordinates': None, 'country': 'SE', 'country_code': None, 'country_concept_id': None, 'municipality': 'Malmö', 'municipality_code': '1280', 'municipality_concept_id': None, 'postcode': '20580', 'region': 'Skåne län', 'region_code': '12', 'region_concept_id': None, 'street_address': 'August Palms plats 1'}

Table 6: *Example ad from 2020. In total 33 variables are presented. Properties nested within a variable are indicated by curly brackets ({}). All names discussed in the variable selection section are consistent with the names in this ad with the exception of occupational group code which is under* legacy_ams_taxonomy_id. *Some values have been purposefully altered to fit the table indicated by '...'.*