JOLTS Overview

Introduction

The Job Openings and Labor Turnover Survey (JOLTS) tells us how many job openings there are each month, how many workers were hired, how many quit their job, how many were laid off, and how many experienced other separations (which includes worker deaths).

The JOLTS survey design is a stratified random sample of 20,700 nonfarm business and government establishments. The sample is stratified by ownership, region, industry sector, and establishment size class. The establishments are drawn from a universe of over 9.4 million establishments compiled by the Quarterly Census of Employment and Wages (QCEW) program which includes all employers subject to state unemployment insurance laws and federal agencies subject to the Unemployment Compensation for Federal Employees program.

Employment estimates are benchmarked, or ratio adjusted, monthly to the strike-adjusted employment estimates of the Current Employment Statistics (CES) survey. A ratio of CES to JOLTS employment is used to adjust the levels for all other JOLTS data elements.

JOLTS data provide information on all pieces that go into the net change in the number of jobs. These components include hires, layoffs, voluntary quits, and other job separations (which includes retirements and worker deaths). Putting those components together reveals the overall (or net) change. JOLTS data provide information about the end of one month to the end of the next, whereas the monthly employment numbers provide information from the middle of one month to the middle of the next.

The JOLTS estimates also are affected by nonsampling error. Nonsampling error can occur for many reasons including: the failure to include a segment of the population; the inability to obtain data from all units in the sample; the inability or unwillingness of respondents to provide data on a timely basis; mistakes made by respondents; errors made in the collection or processing of the data; and errors from the employment benchmark data used in estimation.[1]

Data

One of the more useful indicators is the job seekers ratio, that is, the ratio of unemployed workers to job openings. On average, there were 10.7 million unemployed workers while there were only 6.5 million job openings. This translates into a job seeker ratio of about 1.6 unemployed workers to every job opening. Another way to think about this: for every

---

[1]Nonsampling error occurs when a sample is surveyed rather than the entire population. Which means that there is a chance that the sample estimates may differ from the true population values they represent. The difference, or sampling error, varies depending on the particular sample selected. This variability is measured by the standard error of the estimate. BLS analysis is generally conducted at the 90-percent level of confidence. That means that there is a 90-percent chance, or level of confidence, that an estimate based on a sample will differ by no more than 1.6 standard errors from the true population value because of sampling error. Sampling error estimates are available at the BLS' website
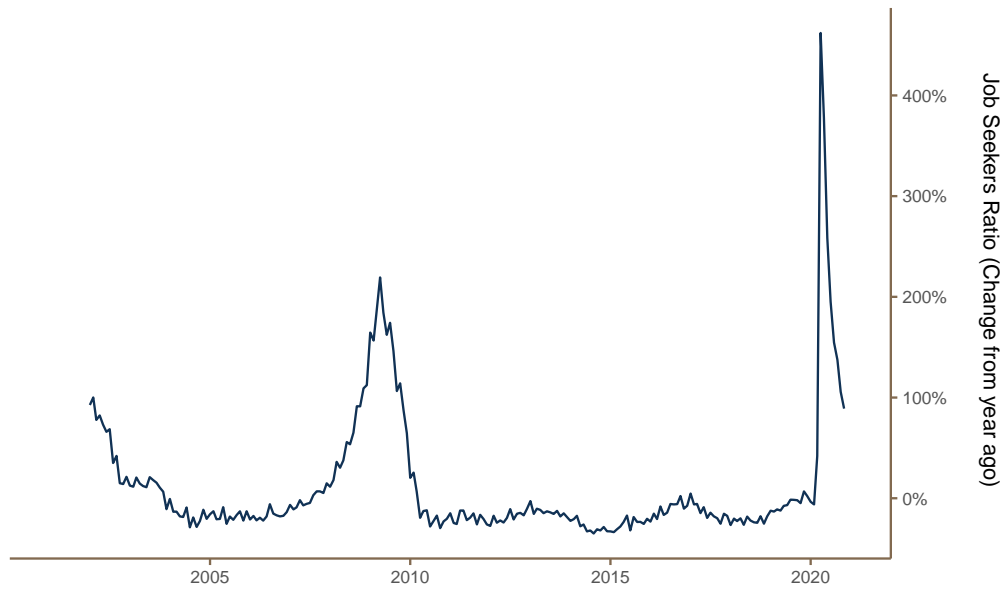
16 workers who were officially counted as unemployed, there were only available jobs for 10 of them.

Job Seeker Ratio
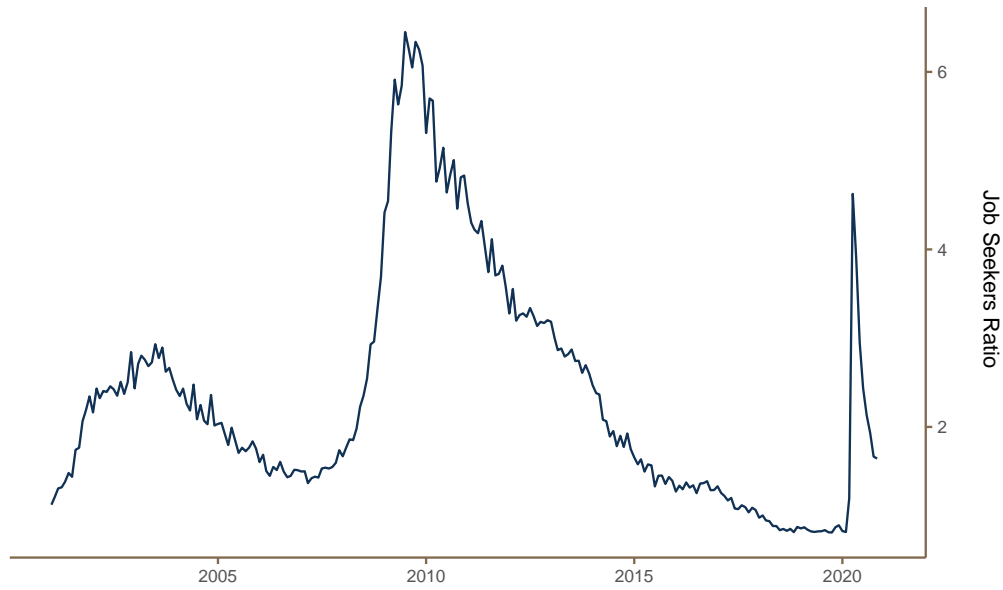
Source: Bureau of Labor Statistics: Job Openings and Labor Turnover Survey
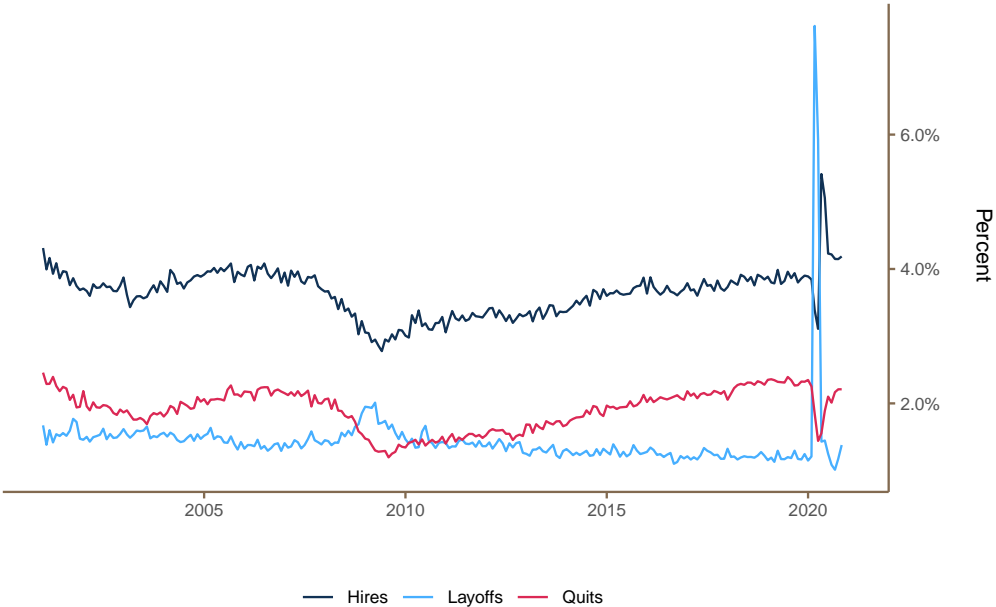
Job Seeker Ratio

Source: Bureau of Labor Statistics: Job Openings and Labor Turnover Survey

Unemployment & Vacany Rate

Source: Bureau of Labor Statistics, Job Openings and Labor Turnover Survey

Hires, Quits, & Layoffs

Source: Bureau of Labor Statistics, Job Openings and Labor Turnover Survey

Hires, Quits, & Layoffs (12−month rolling average)

Source: Bureau of Labor Statistics, Job Openings and Labor Turnover Survey

Hiring Rate (12−month rolling average)

Source: Bureau of Labor Statistics, Job Openings and Labor Turnover Survey

Methods

We produce forecasts using a collection of traditional and non-traditional time series methods. This section provides a general overview of the methods used, their benefits, and their limitations. It cannot be emphasized enough that, no matter the strength of a model, it remains exactly that: a model. As such, all statistical models are "wrong." No matter the method used, any model is an attempt to reproduce ("model") the true data generating process of a data series.

Background

For many familiar with the methodology of time series forecasting, this section presents a brief recapitulation of basic concepts which can be found in any standard textbook such as Montgomery, Jennings and Kulachi (2015) or **?**. For a more advanced treatment, one should see **?** or **?**.

Time series analysis is the procedure of using known data values to fit a time series with a suitable model and estimation of the corresponding parameters. It comprises methods that attempt to understand the nature of the time series.

A major assumption that often provides relief in modeling efforts is the linearity assumption. A linear filter, for example, is a linear operation from one time series x_t to another time series y_t.

$$E[y_t] = L(x_t) = A(L^i)\epsilon_t = \left(\sum_{-\infty}^{\infty} \psi_i L^i \right) \epsilon_t = \mu + \sum_{-\infty}^{\infty} \psi_i \epsilon_{t-i},$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and L^i is the Lag or Backshift Operator defined as,

$$L^i x_t = x_{t-i} \,\forall\, i \in \mathbb{N}$$

the linear filter can be seen as a process that converts the input, x_t, into an output, y_t with a conversion process that involves all (present, past, and future) values of the input in the form of a summation with different "weights", ψ_i, on each observation x_i. Specifically, y_t can be expressed as

$$y_t = \mu + \sum_{i=o}^{\infty} \psi_i \epsilon_{t-i}$$

where the weights are conventionally some form of moving average M. A moving average M of span N assigns weights $\frac{1}{N}$ to the most recent observations such that the estimate can be written as

$$M_t = \frac{1}{N} \sum_{t=T-N-1}^{N} y_t$$

The covariance between y_t and its value at another point in time y_{t+k} is called the auto-covariance at lag k, and is defined by

$$\gamma_k = \text{Cov}(y_t, y_{t+k}) = E\left[(y_t - \mu)(y_{t+k} - \mu) \right]$$

the autocovariance of the series at lag $k = 0$ is simply the variance of the time series. The autocorrelation coefficient at k is then

$$\rho_k = \frac{E[(y_t - \mu)(y_{t+k} - \mu)]}{\sqrt{E[(y_t - \mu)^2]E[(y_{t+k} - \mu)^2]}}$$

$$= \frac{\text{Cov}(y_t, y_{t+k})}{\text{Var}(y_t)}$$

$$= \frac{\gamma_k}{\gamma_0}$$

The collection of these values ρ_k is called the autocorrelation function. For a finite time series, it is necessary to estimate the autocovariance and autocorrelation functions. A usual estimate of the autorcovariance function is

$$c_k = \hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y}), \quad k = 1, 2, 3, \ldots, K$$

and the autocorrelation function is estimated by the sample autocorrelation function (sample ACF)

$$r_k = \hat{\rho}_k = \frac{c_k}{c_o}$$

The stationarity of a time series is related to its statistical properties in time. That is, in the more strict sense, a stationary time series exhibits similar "statistical behavior" in time and this is often characterized as a constant probability distribution in time. However, it is usually satisfactory to consider the first two moments of the time series and define stationarity (or weak stationarity) as follows: (1) the expected value of the time series does not depend on time and (2) the autocovariance function defined as $\text{Cov}(y_t, y_{t+k})$ for any lag k is only a function of k and not time: that is, $y(k) = \text{Cov}(y_t, y_{t+k})$. Weak stationarity, then, implies that the mean of the series does not change with time t and that the autocovariance function, $\gamma(k, t)$, depends on k and t only through their difference $k - t$.

For a time-invariant and stable linear filter a stationary input time series $y(t)$ can be written more succinctly as,

$$y(t) = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_t$$

$$\mu + \sum_{i=0}^{\infty} \psi_i L^i \epsilon_t$$

$$\mu + \Psi(L)\epsilon_t$$

where

$$\Psi(L)\epsilon_t = \sum_{i=0}^{\infty} \psi_i L^i$$

The infinite moving average model has a covariance function in the form,

$$\text{Cov}(y_t, y_{t+k}) = \gamma_y(k, t) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_i \psi_j \gamma_x(i - j + k)$$

The autocovariance function of y_t is then,

$$\gamma_y(k, t) = \sum_{i=o}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \gamma_\epsilon(i - j + k)$$

$$= \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$$

This representation comes from a theorem by Wold and which essentially states that any nondeterministic weakly stationary time series y_t can be represented as in the above manner. A more intuitive interpretation of this theorem is that a stationary time series can be seen as the weighted sum of the present and past random "disturbances."

Although very powerful in providing a general representation of any stationary time series, the infinite moving average model given in is not very useful in practice except for certain special cases.

In finite order moving average or MA models, weights that are not set to 0 are represented by the Greek letter θ with a minus sign in front.

$$y_t = \mu + \left(1 - \sum_{i=1}^{q} \theta_i L^i\right) \epsilon_i$$

$$= \mu + \Theta(L)\epsilon_t$$

where $\Theta(L)\epsilon_t = 1 - \sum_{i=1}^{q} \theta_i L^i$

An interpretation of the finite order MA processes is that at any given time, of the infinitely many past disturbances, only a finite number of the disturbances "contribute" to the current value of the time series and that the time window of the contributors "moves" in time, making the "oldest" disturbance obsolete for the next observation. It is indeed not too far fetched to think that some processes might have these intrinsic dynamics. However, for some others, we may be required to consider the "lingering" contributions of the disturbances that happened back in the past. This will of course bring us back to square one in terms of our efforts in estimating infinitely many weights. Another solution to this problem is through autoregressive models in which the infinitely many weights are assumed to follow a distinct pattern and can be successfully represented with only a handful of parameters. We shall now consider some special cases of autoregressive processes.

ARMA

Autoregressive models are based on the idea that current value of the series, x_t can be explained as a linear combination of past values, $x_{t-1}, x_{t-2}, \ldots, x_{t-p}$, together with some random error ϵ. In other words, a series x_t is linearly dependent upon its past values, the degree to which depends on the lag order p.

Let us first consider again the time series

$$y(t) = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_t$$

$$\mu + \sum_{i=0}^{\infty} \psi_i L^i \epsilon_t$$

$$\mu + \Psi(L)\epsilon_t$$

One approach to modeling this time series is to assume that the contributions of the disturbances that are way in the past should be small compared to the more recent disturbances that the process has experienced. Since the disturbances are independently and identically distributed random variables, we can simply assume a set of infinitely many weights in descending magnitudes reflecting the diminishing magnitudes of contributions of the disturbances in the past.

A simple and yet intuitive set of such weights can be created following an exponential decay pattern. For that we will set $\psi_i = \phi^i$, where $|\phi| < 1$ to guarantee the exponential "decay." In this notation, the weights on the disturbances starting from the current disturbance and going back in past will be $1, \phi^2, \phi^3, \ldots$

$$y_t = \mu + \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots$$

$$= \mu + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}$$

We also have

$$y_{t-1} = \mu + \epsilon_{t-1} + \phi\epsilon_{t-2} + \phi^2\epsilon_{t-3} + \cdots$$

$$= \mu + \phi\left(\sum_{i=0}^{\infty} \phi^i \epsilon_{t-1-k} + \epsilon_t\right)$$

combining these two equations together we can represent an AR(1) model as a linear process given by

$$\mu + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i} = \mu + \phi\left(\sum_{i=0}^{\infty} \phi^i \epsilon_{t-i} + \epsilon_t\right)$$

$$\mu + \epsilon_t + \sum_{i=0}^{\infty} \phi^i y_{t-i} + \phi\mu$$

$$\mu - \phi\mu + \sum_{i=0}^{\infty} \phi^i y_{t-i} + \epsilon_t$$

$$\delta + \sum_{i=0}^{\infty} \phi^i y_{t-i} + \epsilon_t$$

where $\delta = (1-\phi)\mu$. A more general way of writing an autoregressive procress is through operator notation as

$$\epsilon_t = \left(1 - \sum_{i=1}^{i} \phi_i L^i\right) y_t$$

$$= \Phi(L)y_t$$

where $\Phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$

The above process in is called a first-order autoregressive process, AR(l), because it can be seen as a regression of y_t on y_{t-1} and hence the term autoregressive process.

Autoregressive–moving-average (ARMA) models provide a parsimonious description of a (weakly) stationary stochastic process in terms of two polynomials, one for the autoregression (AR) and the second for the moving average (MA). The general ARMA model was described in the 1951 dissertation of Peter Whittle, Hypothesis Testing in Time Series, which generalized Wold's autoregressive representation theorem for univariate stationary processes to multivariate processes, and was popularized in by Box and Jenkins in their classic book **?**.

$$y_t = \delta + \sum_{i=0}^{p} \phi^i y_{t-i} + \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{i-1}$$

$$\Psi(L)y_t = \delta + \Theta(L)\epsilon_t$$

which has an infinite moving average process $\mathrm{MA}(q)$

$$y_t = \mu + \sum_{i=o}^{\infty} \psi_i \epsilon_{t-i}$$

ARIMA

The Autoregressive Integrated Moving Average (ARIMA) model is a generalization of an Autoregressive Moving Average (ARMA) model to overcome the possible violation of the assumption that the series x_t is stationary. In order for inferences drawn from the estimates to be reliable when the data are not stationary, the data must be differenced some number of times d to remove seasonality and trend. If a stochastic process has to be differenced d times to reach stationarity, it is said to be integrated of order d or $I(d)$.

A standard convention for expressing the ARIMA model is $\mathrm{ARIMA}(p, d, q)$, where:

- p denotes the number of lag observations included in the model, also called the lag order.

- d denotes the number of times that the series is differenced, also called the degree of differencing.

- q denotes the size of the moving average window.

We will call a time series homogeneous, non-stationary if it is not stationary but its first difference, that is, $w_t = y_t - Y_{t-1} = (1-L)y_t$, or higher-order differences, $w_t = (I-L)^d y_t$ produce a stationary time series. We will further call y_t an autoregressive integrated moving average (ARIMA) process of orders p, d, and q ARIMA(p, d, q) if its d-th difference, denoted by $w_t = (I-L)^d y_t$ produces a stationary ARMA(p, q) process. The term integrated is used since, $d = 1$, for example, we can write y_t as the sum (or "integral") of the w_t process as

$$y_t = \delta + \sum_{i=0}^{p} \phi^i y_{t-i} + \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{i-1}$$

$$\left(1 - \sum_{i=0}^{p} \phi^i L^i\right) y_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \epsilon_t$$

$$\left(1 - \sum_{i=0}^{p} \phi^i L^i\right) = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)(1-L)^d$$

$$\left(1 - \sum_{i=0}^{p} \phi^i L^i\right)(1-L)^d y_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \epsilon_t$$

$$\Phi(L)(1-L)^d y_t = \delta + \Theta(L)\epsilon_t$$

Selection of the hyperparameters (p, d, q) are often chosen through an inspection of the ACF and PACF, but can also be selected through AIC (Akaike Information Criterion), AICc (corrected AIC) and BIC (Bayesian Information Criterion). But note that the selection of the hyperparameters is not unique.

SVM

RNN

Random Forest

Unit Root

A unit root is a feature of some stochastic processes (such as random walks) that can cause problems in statistical inference involving time series models. A linear stochastic process has a unit root if 1 is a root of the process's characteristic equation. Shocks to a unit root process have permanent effects which do not decay as they would if the process were stationary. The characteristic roots (roots of the characteristic equation) also provide qualitative information about the behavior of the variable whose evolution is described by the dynamic equation. For a differential equation parameterized on time, the variable's evolution is stable if and only if the real part of each root is negative. For difference equations, such as a standard time series, there is stability if and only if the absolute value of each root is less than 1.

An augmented Dickey–Fuller test (ADF) tests the null hypothesis that a unit root is present in a time series sample. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity.

The intuition behind the test is that if the series is characterised by a unit root process then the lagged level of the series (y_{t-1}) will provide no relevant information in predicting the change in y_t besides the one obtained in the lagged changes. The OLS estimate (based on an n-observation time series) of the autocorrelation parameter ρ is given by

$$\hat{\rho}(n) = \frac{\sum\limits_{t=1}^{n} y_{t-1} y_t}{\sum\limits_{t=1}^{n} y_t^2}$$

if $|\rho| < 1$ then

$$\sqrt{n}\,(\hat{\rho} - \rho) \implies \sim N(0, 1 - \rho^2)$$

To compute the test statistics, we fit the augmented Dickey–Fuller regression

$$\Delta y_t = \alpha + \beta y_{t-1} + \delta + \sum_{j=1}^{k} \psi_i \Delta y_{t-i} + \epsilon_t$$

Depending on the specifications, the constant term α or time trend δ is omitted and k is the number of lags specified.