

The Conference Board Help Wanted OnLine® Data Series Technical Notes

Methodology

Data Collection. The HWOL program uses data collected from over 16,000 online job sources including traditional job boards, corporate boards, and social media sites. Internet job sources that are aggregators (i.e. only scrape ads from other sources and provide no unique ads) are identified and removed from active collection in order to eliminate a major source of duplication in counting online ads. New job sources are identified using independent research and recommendations from industry sources across the U.S. This process results in periodic updates to the HWOL coverage. Job sources that cover smaller niche markets are also included in HWOL; however, smaller local job sources in an area with a limited number of ads may not be targeted for collection.

Duplication. Data in the HWOL time series reflect unduplicated ads. A major issue in producing estimates of the actual level of advertised vacancies for a geographic area is the elimination of duplicate ads. There is a significant amount of ad-scraping across job sources, and there are large nationwide job boards that contain only scraped ads. As noted above, the HWOL program first identifies job sources that are only aggregators of ads from job sources and eliminates these from active collection. Ads are then unduplicated across all in-scope HWOL job sources, and duplicates are eliminated from the HWOL published estimates. This process significantly limits the level of potential duplicates in the final estimates. The unduplication process reduces the count of overall ads collected from over 13 million ads to over 4.5 million ads after unduplication and therefore indicates that duplicates represented about 2 out of 3 of the ads prior to unduplication. The resultant unduplicated ad levels for the HWOL program compare favorably to those produced by the BLS Job Openings Labor Turnover (JOLTS) program after allowing for coverage and definitional differences.⁴ © 2013 The Conference Board. All rights reserved.

Occupational coding. The HWOL program uses the Federal Government's Standard Occupational Classification (SOC) Manual to assign an occupational code to each ad. Occupational coding is done at the 6-digit Standard Occupational Classification (SOC) level and the 8-digit O*Net level using autocoder software selected by The Conference Board for its accuracy. The Conference Board periodically updates the O*Net classification of HWOL data and reclassifies the entire HWOL database of ads with each new Federal Government revision to the SOC manual.

Area coding. The area coding for an ad is determined first by the location cited in the text of the ad itself; approximately 93% of all ads are coded to a county/city level. Of the remaining ads, approximately 5% are coded as "Statewide" with 2% coded as "nationwide" ads; nationwide ads would appear in the national total but not in any regional, state, or metropolitan area totals.

Estimation. All HWOL level counts are produced as a direct sum of the unduplicated city counts (i.e. a county total is a sum of the cities contained within the county). There are 2 exceptions: the State counts, which also include "Statewide ads" (i.e. ads which are only posted at the State level -- approximately 2% of total ads), and the US counts, which include both Statewide ads along with "Nationwide ads" (i.e. ads which are only posted at the US level -- approximately 5% of total ads).

Seasonal Adjustment. The HWOL program uses the Census Bureau's X-12 seasonal adjustment software to annually update the seasonal adjustment factors for each of the publication time series. The new seasonally adjusted series are released with the publication of each year's January data.

Reliability of Estimates

The HWOL program is basically a universe count and is not subject to the typical sampling error and non-response error components associated with most statistical surveys. The non-sampling error sources for the HWOL program would include population under-coverage due to missing a portion of the targeted population (e.g. a large Internet job source) and over-coverage due to the inability to fully eliminate duplicate ads from survey estimates. Additional potential sources of non-sampling error would include occupational and/or geographic coding errors which could affect the proper classification of individual ads.

Source

The Conference Board Help Wanted Online Data Series® (HWOL)