Emsi Occupation Employment and Earnings Methodology Update: **OES Time Series**

Introduction

Emsi has made two improvements in how we use OES. These improvements allow us to drastically improve the quality of occupation employment and earnings data that we deliver to our clients.

First, Emsi now offers historical occupation earnings data back to 2005. Because the BLS advises against combining multiple years of OES to create a time series (a multi-year dataset) of employment or earnings, Emsi has only ever provided earnings from the latest year of OES available.

Emsi has always provided historical occupation employment counts as a time series because our employment counts are based on the BLS's QCEW dataset, which is a time series by definition. In the past, we used OES's latest staffing patterns and regional employment estimates, back-projected for older years, to transform QCEW-based industry job counts into occupation job counts. The second improvement to Emsi occupation data is that we now use historical OES staffing patterns and regional employment data to transform industry data, rather than just using back-projected, current-year OES.

The first part of this paper outlines the methodology Emsi now uses to create occupation employment and earnings data, making use of OES as a time series. The second section addresses how Emsi handled the BLS's cautions about treating OES as a time series.

Timeline of Changes

- Prior to the 2018.3 datarun, Emsi only made use of current-year OES data to create regional staffing patterns and occupation employment and earnings data. The current-year OES national staffing pattern was projected back to 2001 and forward ten years using NIOEM and Emsi industry data, then regionalized using current-year regional OES occupation data, also projected backward and forward. Occupation earnings were only estimated for the current year.
- With the 2018.3 datarun, Emsi began making use of historical OES national staffing patterns back to 2005. However, the regionalization of those national OES staffing patterns was still being performed with current-year OES regional occupation data, projected backwards and forwards for historical data and projections, respectively.
- With the 2019.1 datarun, Emsi continued to research and laid the groundwork for future integrations of historical OES data.
    - First, Emsi tightened constraints on occupation employment estimates for undisclosed OES data, to cluster them more tightly around disclosed OES data (Emsi's previous unsuppression methodology didn't match disclosed OES data as closely as it could have).
    - Second, Emsi improved the algorithms that estimate top-coded earnings data points (data points that OES suppresses because they are above $100/hour).
    - Third, Emsi began applying proprietary occupation percentile earnings aggregation methodology to the earnings unsuppression process. On the whole, percentile wage curves became more reasonable and much more statistically defensible.
- With the 2019.3 datarun, Emsi produced historical occupation earnings estimates. Prior to this datarun, Emsi only produced current-year occupation earnings. Additionally, latest-year earnings now take advantage of the OES time series data, meaning that the unsuppression of earnings data is informed by other years in the dataset, resulting in a more cohesive year-over-year set.

Emsi Occupation Employment and Earnings Methodology

IMPORTING AND CLEANING OES DATA

The first step in creating occupation employment and earnings data is cleaning the source OES data. Because the BLS did not design OES to be a time series, it is necessary to do some preliminary cleaning to get the data into a state where it can be used as a time series. Data cleaning involves a number of steps including uncovering and accounting for missing data that was suppressed after its release because errors were found in it; adjusting all years of OES so the data file structures, headers, and codes are the same; and filtering out duplicate data and certain mid-level aggregations that can't be used.

Emsi's cleaning of OES data makes two methodological assumptions. First, for the handful of occupations for which OES only produces annual earnings (e.g. teachers and airline pilots), Emsi converts to hourly earnings by dividing the annual figure by 2,080 (number of hours worked in the average full-time year). Second, Emsi ignores data for New England City and Town Areas (NECTAs) because they do not align with county boundaries and thus are incompatible with Emsi geographical hierarchies. Emsi instead uses state-level data for the six New England states (New Hampshire, Vermont, Massachusetts, Connecticut, Rhode Island, and Maine).

At the end of each part of the importing and cleaning process, Emsi tests data definitions and metadata against the raw OES data, and also ensures that employment and average earnings continue to aggregate to higher classification levels (e.g. 5-digit SOC figures sum to their parent 2-digit SOC figures) correctly where the data is disclosed.

SEEDING OES DATA

Seeding is the process of coming up with initial estimates for suppressed data points. Seeding cannot be arbitrary or random, because the seed values selected will affect the final outcome of the unsuppression. Seeds are collected and calculated in many different ways depending on the type of data being seeded. We first seed industry and sector totals, then move to staffing patterns, and finally to regional employment data.

QCEW SEEDING OF OES INDUSTRY AND SECTOR TOTALS

The first step is to seed industry and sector totals with data from QCEW. QCEW is ideal because OES benchmarks to monthly QCEW (from May of the OES year and November of the previous year). Suppressed OES data are simply seeded with the corresponding QCEW value whenever the QCEW value is disclosed.

YEAR-OVER-YEAR STAFFING PATTERN SEEDING

Next, undisclosed OES national staffing data is seeded using a staffing pattern from the subsequent year of OES multiplied by current-year OES occupation totals. This approach reduces volatility in the time series by making each year's staffing pattern slightly more like the subsequent year, while still remaining consistent with current-year estimates. For the latest year, which does not have a seed staffing pattern (staffing patterns are seeded in reverse chronological order, with 2017 seeding 2016 and so on), Emsi draws on OES data at higher levels of aggregation to provide seed values.

Year-over-year staffing pattern seeding also allows us to handle cases where OES moves employment to different industries or occupations. For example, prior to 2014, OES categorized local government gambling establishments and casino hotels in NAICS 999300 (Local Government). Beginning in 2014, OES moved these establishments to NAICS 7132 and 72112, respectively. In order to create a consistent time series, we need to move local government gambling from 9993 to 7132 and 72112 for all years prior to 2014. This is achieved by using the seeds from the subsequent year to create a ratio of gambling to total local government. That ratio is then applied to NAICS 999300 in years prior to 2014 to determine how much employment to move to the gambling industries.

MULTI-YEAR SEEDING OF REGIONAL EMPLOYMENT DATA

After staffing patterns are fully seeded, regional employment data is seeded. Where possible, disclosed data from previous and subsequent OES years are used to seed regional employment. As we will discuss later in relation to BLS cautions about OES time series, this seeding method is preferred since it reduces year-over-year volatility in the OES time series by seeding each year with all other years.

The underlying assumption is that if data exists for an occupation-geography combination in one year, it is likely that it should also exist in a different year. If data exists for the occupation-region combination in question in either prior or subsequent years, we make the data look more like the year being seeded. This is done by applying the difference at the national level between the year being seeded and the year for which a seed was found for the occupation in question. The national-level change is applied to the regional seed, similar to a unit conversion equation:

target seed value = other year local seed * (target year national disclosed figure / corresponding other year national disclosed figure)

For instance, this example provides a seed for Registered Nurses for 2016, using a value from 2014:

2016 RN seed estimate = 2014 RN value * (2016 national disclosed figure for RNs / 2014 national disclosed figure for RNs)

If we have data for the region in both prior and subsequent years, we perform the above adjustment function on both available data points, then average the two to create our final seed.

Any employment seeds filled with these "multi-year" seeds are adjusted so they do not sum to more than disclosed summary data within the year being seeded. It is important that we make this adjustment before proportioning (unsuppressing), because allowing seeds to be too large will reduce the quality of other seeds that will be added later.

This multi-year seeding method is used to unsuppress earnings as well as employment.

QCEW SEEDING OF REGIONAL EMPLOYMENT DATA

Seeds are also created for OES regional employment data by running regional QCEW (formatted like OES) through the national unsuppressed OES staffing pattern, resulting in occupation employment seeds based on the national staffing pattern. The underlying assumption here is that regional staffing patterns will mimic national staffing patterns. Obviously this is not a valid assumption for final data, but it is a safe and reasonable assumption to use in creating seed values that will be used to help unsuppress the regional data. Using QCEW in combination with the national staffing pattern will tend to increase the cohesion of the entire dataset across geographies while maintaining the regional distinctives disclosed in each region's OES occupation totals.

UNSUPPRESSING OES DATA

Once quality seeds have been selected for each year of OES data, Emsi runs proprietary unsuppression algorithms (called "proportioners") that iteratively and proportionally adjust the data until detailed data correctly sums to summary data. Unsuppression is always done on each OES year individually (before we create a time series) to ensure that each year of OES data is complete and internally consistent. After each year is unsuppressed individually, the time series will be created.

The unsuppression process consists of running bi-proportional algorithms on the seeded data until the detailed data correctly sums to summary data. The process can be thought of as a matrix with the dimensions along the rows and columns.

	Idaho MSA 1	**Idaho MSA 2**	**Idaho MSA 3**	**Idaho MSA 4**	**Disclosed SOC sums**
SOC 1	-	820	850	-	2130
SOC 2	780	-	-	930	2700
SOC 3	640	620	370	440	2070
SOC 4	460	240	-	-	1150
Disclosed MSA sums	-	-	1630	1530	8050

Disclosed values are fixed and may not change. Seed values (represented above by dashes) are allowed to change so that the rows and columns can sum to known employment totals in the columns for MSAs and known employment totals in the rows for SOCs. The bi-pro works iteratively within the matrix, adjusting the seed values so that first the rows sum. At this point, columns will not sum. The seeds are then re-adjusted so that the columns sum. This causes the rows to no longer sum. Iterating through the column and row summing once constitutes one “pass” of the bi-pro. As passes are repeated and seed values are constantly re-adjusted, row and column sums gradually come into alignment with each other. Unsuppression is achieved when the matrix is balanced, that is, column and row sums equal each other.

EARNINGS ESTIMATION

After unsuppressing employment, we move on to earnings unsuppression and estimation.

MINIMUM WAGE FLOORS

The first step in unsuppressing earnings is to set minimum wage bounds. Emsi uses same-year federal/state minimum wage data to create the lower bound for wages to match OES methodology: “We adjust the aged lower bounds of the older panel survey data to be at least the current federal minimum wage, or the state minimum wage if higher.” Minimum wage is used to constrain the lower bound on seeds throughout the rest of the process to prevent skewing estimates lower than they should be.

AVERAGE EARNINGS

Average earnings are unsuppressed first. Emsi uses unsuppresed OES regional occupation employment to unsuppress average earnings, since average earnings can be weighted by employment and aggregated fairly well. Average earnings are unsuppressed similarly to employment--they are seeded and then run through a bi-pro.

PERCENTILE EARNINGS

Percentiles cannot be aggregated or distributed (proportioned), which means that the bi-pro algorithm used to unsuppress employment and average earnings cannot be used to unsuppress percentile earnings. Instead, the methodology for estimating percentiles is akin to a nearest neighbor algorithm that takes into account unsuppressed average earnings and percentile earnings seed values drawn from other OES years.

Percentile earnings data are seeded with multi-year seeds drawn from other years of OES data. Emsi also uses unsuppressed average earnings to inform the unsuppression of percentile earnings. Great care is taken to keep percentiles in agreement with average earnings by making incoming multi-year percentile seeds consistent. Consistency is achieved by adjusting percentiles to match average earnings after seeding and then matching again after unsuppression, and estimating any remaining undisclosed percentiles using average earnings.

Emsi uses an Alternative Minimum Mean and Alternative Maximum Mean (AMMs) to determine how consistent average earnings are with the percentiles. The alternative minimum mean works by assuming that all of the people within a certain percentile make the least amount possible (e.g. everyone in the 10th percentile to 25th percentile range makes exactly 10th percentile wage), and conversely with the alternative maximum mean (e.g. everyone in the 10th percentile to 25th percentile range makes exactly 25th percentile wage). Unless the values are disclosed, forcing them to remain as they are, the percentiles and/or the average are adjusted to fall within the AMMs.

UNSUPPRESSING TOP-CODED VALUES

One of the ways OES suppresses earnings is by "top-coding" earnings at $100/hr (for 2018; the top-code value changes over time). By definition, we can be sure that any top-coded value is not less than $100/hr. Fortunately, OES only suppresses average earnings as top-coded if all percentiles are top-coded. When at least one percentile is disclosed, average earnings are disclosed even if the average earnings value is greater than the top-code value. Most average earnings that are greater than the top-code value are disclosed because at least one of the lowest percentiles is also disclosed. Given the amount of data that was therefore disclosed despite being above the top-code value, Emsi was able to develop an algebra-based method of estimating top-coded values that only relies on the data available in OES.

FINAL EARNINGS TESTS

Final earnings data is tested to ensure that certain methodological criteria are met. These tests include:

- The number of AMM disagreements between average earnings and the percentiles must be reasonable (<0.5%).
- Percentiles must ascend (we do allow flat percentiles for small samples)
- 10th percentile wages must not be lower than minimum wage
- 90th percentile wages must be reasonable (<$1000/hr)
- Earnings match disclosed OES data where possible
- Top-coded earnings do not fall below the top-code value ($100 in 2018)

CREATING A TIME SERIES

After each individual year of OES has been unsuppressed, the next steps in the process combine OES years to create a time series.

TIME SERIES DIMENSION STANDARDIZATION

The first step in creating a time series of OES data is to standardize the dimension hierarchy definitions. Over the years, OES has made many changes to all three of the primary dimensions: occupation, industry/sector, and geography. A time series requires that each dimension use only one hierarchy definition (one version of NAICS, for example) across all years; therefore older years of data must be re-calculated in terms of the latest hierarchies.

Dimension standardization is a standard Emsi procedure, and the process is the same for OES data. Emsi begins by creating relationship files between the different definitions that exist for each hierarchy, then creates groupings of related codes and runs then through a proportioning algorithm that calculates the optimal breakout for each data point in the source data in terms of the latest hierarchy definition. This process is repeated for all dimensions on all years of data until the entire time series uses a consistent set of the latest definitions.

TIME SERIES OUTLIER REMOVAL (STAFFING PATTERNS)

Each OES release has anywhere between 15 and 30 occupations (2-4%) with questionable shifts in employment between the latest and the prior year that needed to be investigated. In a given release, roughly 40% of these questionable shifts and the outliers they cause can be attributed to hidden methodology changes.

Emsi uses a complex outlier removal algorithm to correct for changes in the survey that cause shifts in OES staffing patterns and occupation data. The algorithm detects outliers in occupation totals between years and corrects for them, and is applied throughout the time series in reverse order so the latest year of OES remains unchanged, and previous years are adjusted to more closely reflect more recent data.

Known, legitimate changes in industry employment can cause false positives for outlier detection. In order to prevent these legitimate outliers from being corrected, Emsi first estimates a year's expected occupation totals using the adjacent year's staffing pattern multiplied by the year in question's industry/sector totals. This estimate yields occupation totals we'd expect to see given current QCEW industry data combined with the adjacent year's staffing pattern. Occupation totals using both the staffing pattern and the industry data from the year in question can be compared to the expected values to identify outliers caused by methodology shifts in the OES staffing pattern for the year in question.

Occupation totals with outliers removed are inserted back into the dataset and the dataset is proportioned to keep industry and occupation totals consistent. This forces the current staffing pattern to be much more similar to the subsequent staffing pattern that was used to detect outliers, effectively reducing the noise introduced by undocumented changes in the OES survey methodology.

Data prior to 2005 is built by back-projecting the time series using all available years to fill in years 2004-2001. Significant changes to OES were made in 2002 that render data prior to 2005 too difficult to work with. Since OES is a rolling 3-year survey, the 2002 changes also affected 2003 and 2004 data. Therefore, Emsi back-projects data for 2001-2004.

TIME SERIES SMOOTHING (REGIONAL OCCUPATION EMPLOYMENT DATA)

Emsi employs a Reverse Kaufman Adaptive Moving Average (KAMA) in order to reduce time series volatility in regional occupation employment trends. The algorithm is designed to aggressively smooth when OES employment is volatile, but leave non-volatile trends largely unchanged. KAMA was originally designed to model stock market trends while remaining stable during brief, meaningless spikes and dips in the market. Emsi applies this concept to labor market data to achieve the same effect with employment trends--KAMA keeps trends stable through brief, meaningless spikes and dips in the labor market, but follows given trends closely when those trends are relatively stable.

After smoothing employment using KAMA, Emsi controls the smoothed regional occupation employment data to national employment totals built from the national staffing pattern. This brings the regional employment data back into agreement with national data and ensures proper aggregation from regional totals to national totals.

At this point, creation of the employment time series is complete. All years use the most current classification hierarchies and have been brought into alignment with one another. At this point, earnings data is still compartmentalized by year. The next step is to convert earnings data to a time series as well.

RECALCULATING EARNINGS AFTER TIME SERIES CREATION

Initially, earnings are calculated within each OES year. After the employment time series is built, earnings are recalculated as a time series using time series employment to inform the recalculation. The recalculation step is necessary; without it, our earnings are simply earnings data for standalone OES years, which is what the BLS warns against doing with OES.

The recalculation process combines the new OES employment data in time series format, which uses the current OES classification structures, with OES earnings in each year's original classification structure. We convert each year's average and percentile earnings into the current OES classification structures using Emsi's proprietary occupation earnings aggregation methodology. In this way, the latest year of OES is treated as ground truth, and earlier years are brought into alignment with its industry, occupation, and geographical dimensions.

To bring older OES earnings data forward into current classification hierarchies, Emsi first combines all the forward mappings from the source year's classifications to the current year's classifications to create a single mapping between the two years. That mapping is then used to determine which codes have changed over the intervening years and also how heavily to weight the source earnings estimates that we'll be combining to create the final earnings estimates. Emsi then recalculates earnings using all the data in the source year and source classification that influenced the earnings in the current year and current classification. If a single datapoint influenced the final data point, earnings are simply copied to the final dataset. This happens fairly often. Finally, Emsi ensures that employment did not change and that all earnings have been estimated.

COMPLETION

At this point, OES employment and earnings data have both been converted to a time series and the process is complete.

Next, we discuss the BLS's cautions about treating OES as a time series.

Navigation of BLS Cautions Against OES Time Series

The BLS advises against using historical OES data as a time series for the reasons outlined below. For each caution, we outline our approach to the problem and how we mitigate the risk(s) presented by the BLS.

Simply, their cautions are the following:

- OES data over time contains industry, occupation, and geography classification shifts.
- In 2002, the survey reference period was changed and methodology for estimating mean wages was changed, creating a rift between data for 2002 and prior years, and 2003 and later years.
- Certain permanent features of OES methodology, such as "bringing forward" earnings data in older survey panels within an OES year, or using six survey panels, may mask sudden changes in employment or earnings.
- OES data collection procedures have changed over time, and these changes can result in large shifts in data that are not formally documented and are therefore invisible to users of OES data.

For more information on each of these cautions, see the BLS's own thorough treatment of the subject.

Each of these cautions is addressed in the following section.

Changes in industrial, occupation, and geographical classifications

Emsi deals with industry, occupation, and geographical classification changes in virtually every data source we work with. Any LMI provider offering more than a few years of historical employment or earnings data will be well-accustomed to dealing with classification changes, as this is a common issue across most government datasets.

Emsi creates mappings between old and new classifications using proportioning algorithms. These algorithms determine the relationship between the old and new classifications and, where necessary, determine the optimal proportioning needed to transform data in an old classification into a new classification. For instance, in the case of one SOC code in an old hierarchy splitting into two codes in a new hierarchy, proportioning algorithms determine the relationship between the old SOC code and the new ones, and map data out of the old hierarchy and into the new one.

Changes in the Survey Reference Period and Changes in Mean Wage Estimation Methodology

Two of the BLS's cautions against using OES as a time series relate to changes made to the survey in 2002.

In 2002, OES switched from surveying from October through December to surveying in May and November. Variations in seasonal employment at different times during the year caused large shifts in employment during the year in which this change was introduced.

Also in 2002, OES made changes to how average wages are calculated for some higher-paying occupations. The changes resulted in higher average wages for a handful of occupations.

Since each year's release of OES contains data collected over the previous three years, the seasonality and older method of calculating average wages for higher-paying occupations remained in the data through the 2004 release. To skirt these problems, Emsi does not use OES data prior to 2005, but instead back-projects OES staffing for the years 2001-2004, and Emsi historical occupation wages are only available back to 2005.

PERMANENT FEATURES OF OES METHODOLOGY

The BLS also states that the use of the six-panel survey method results in large, sudden changes in actual employment as reported in individual panels being smoothed out over time, as an initial large change to employment is damped by the other five panels already in the survey set. Emsi submits that this is a positive feature of OES, which is volatile by nature even when the six bi-annual panels are combined. Anything that is done to lessen volatility makes the survey more usable as a time series.

Additionally, since Emsi relies on QCEW for employment numbers and uses staffing patterns from OES applied to QCEW to create occupation job estimates, changes in employment in the OES data do not affect Emsi occupation job counts, except indirectly, through Emsi's final regionalized staffing patterns. In Emsi data, industry job counts from QCEW carry far more weight in determining occupation job counts than OES does.

CHANGES IN HOW OES DATA IS COLLECTED

Over the years, the BLS has changed the layout of the OES survey, including how occupations are listed to employers. These changes in survey design can affect how employers fill out the survey, causing apparent volatility in employment and earnings figures when the data are analyzed over time. Historically Emsi users have noticed mysterious changes in occupation data each year when Emsi updates to the latest available version of OES. Hidden survey methodology changes are often the culprit in these shifts. Because Emsi used to use each year of OES in isolation, it was impossible to deal with the shifts; we could only note that they had occurred and were probably caused by a change by the BLS in the OES survey methodology. Identifying and dealing with these hidden changes was the most difficult part of making OES into a time series. The techniques described in the process sections above, particularly the smoothing, outlier removal, and multi-year seeding, are techniques that were chosen particularly to help deal with hidden OES survey methodology changes.

HIDDEN METHODOLOGY SHIFTS (STAFFING PATTERNS)

Emsi's new methodology identifies hidden OES methodology changes by comparing actual and expected occupation job counts for each year of OES data. This is done in a reverse-chronological order: for instance, 2017 staffing patterns are applied to 2016 job counts to create 2016 expected job counts, and then those counts are compared to the result of applying 2016 staffing patterns to 2016 job counts. This produces the employment counts that would be generated if the OES staffing pattern were to remain static over 2017 and 2016, and identifies large shifts in employment that are caused by methodology changes in the staffing patterns. By creating expected occupation counts based on the OES staffing pattern from an adjacent year, we are able to pinpoint large shifts in employment data that are caused by hidden methodology shifts in OES staffing patterns between OES years.

Since there is no future year to compare to, latest-year OES data is not adjusted to an expected pattern like prior years' data; rather, the staffing pattern is applied as-is to that year's industry data, and is then used to form the expected pattern for the prior year. Since Emsi's methodology essentially chains years of OES data together using expected and actual breakouts, either the first year in the chain or the last year in the chain must be unaffected by the other years in the chain, and must affect all subsequent years in the chain. We elect to allow the latest year's data inform prior years, rather than allow the oldest data to inform the newest data.

Each OES release has anywhere between 15 and 30 occupations (2-4% of occupations) with questionable shifts in employment that needed to be investigated. In a given release, roughly 40% of the questionable shifts can be attributed to hidden methodology changes in staffing patterns between years.

Emsi's underlying assumption in handling OES shifts is that actual change in the labor market overwhelmingly takes place over time, and not all in the course of one OES update. Methodology changes, on the other hand, are binary, causing instantaneous shifts when implemented. Therefore, generally speaking, methodology changes in OES will present themselves as sudden changes that cannot be explained by or linked to matching industry trends in QCEW, whereas actual changes in the labor market will present themselves more gradually and over time, and will also show up in QCEW.

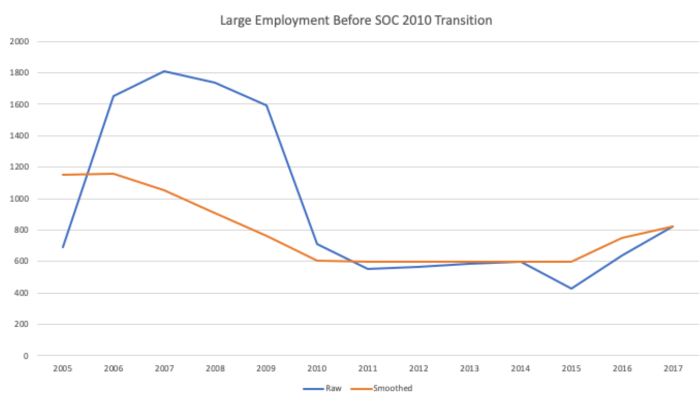
Any occupation with a percent difference greater than 10% between the actual total and expected total patterns is considered an outlier and is corrected. The correction is done by proportionally adjusting the occupation-by-industry totals within the staffing pattern for the occupation. For the most part, the corrections balance each other out (i.e. some outliers are too high; some are too low). Any residual outlier remainders are distributed proportionally across non-outlier occupations. Generally speaking, outliers are more heavily adjusted, while non-outliers receive only very slight adjustments as needed, to absorb residual outlier remainders.

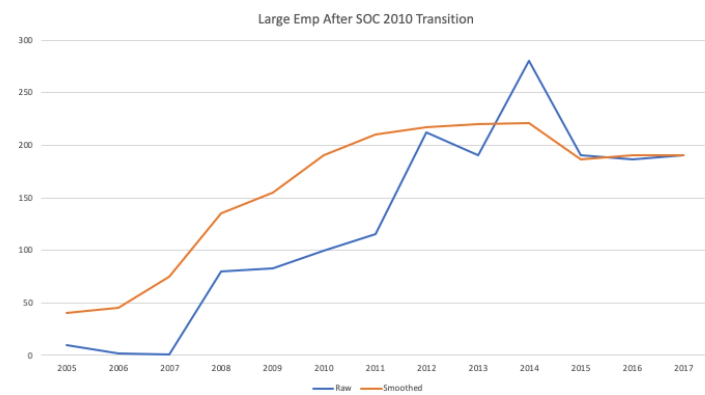
HIDDEN METHODOLOGY SHIFTS (REGIONAL EMPLOYMENT)

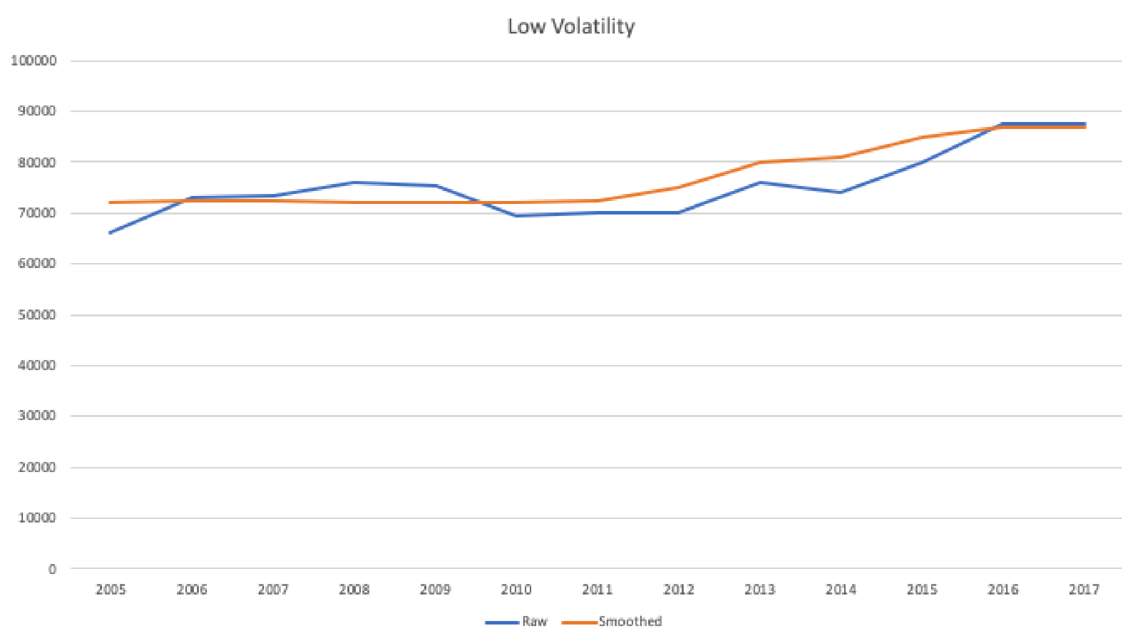
To make use of prior years' regional occupation employment data, Emsi first had to work through the data, dealing with sudden employment shifts caused by the same hidden methodology changes that affected national staffing patterns. The Reverse Kaufman Adaptive Moving Average (KAMA) was designed for this purpose, originally for application to stock market prices, but Emsi has adapted it to perform the same function on employment over time. Because different years of OES can show volatile spikes and dips in employment that are not reflected or confirmed by QCEW industry data, we use KAMA to smooth these shifts and prevent sudden employment changes due to OES methodology changes from heavily affecting the final OES time series.

The below graphs demonstrate KAMA's ability to damp large shifts while remaining close to stable data over time. The first two graphs show KAMA's effects on volatile trend lines in an occupation, largely due to the SOC 2010 transition.

The final graph shows KAMA tracking closely with the stable trend of an occupation largely unaffected by the SOC 2010 transition.



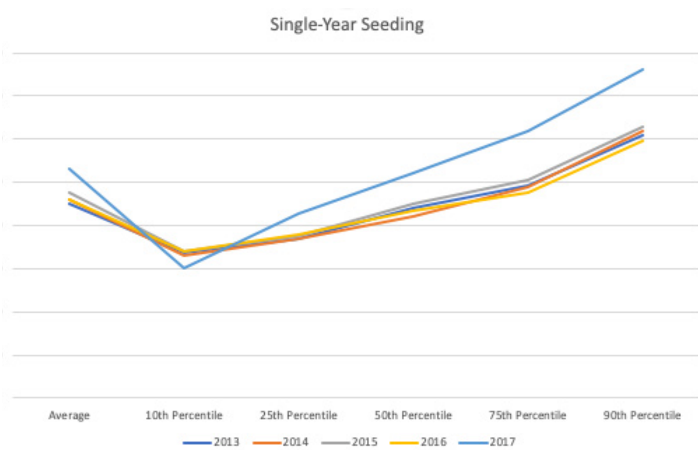




MULTI-YEAR SEEDING

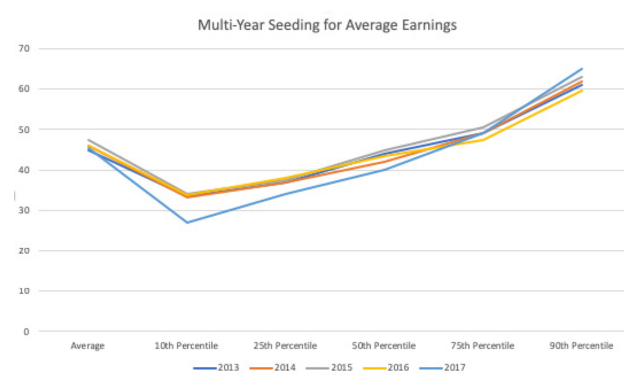
As described above, Emsi uses a multi-year seeding technique to unsuppress OES employment and earnings. Multi-year seeding refers to the practice of using other years of OES data to supply seed values where needed for OES unsuppression. Disclosed data from previous and subsequent years of OES are used as seeds. Multi-year seeding helps deal with hidden OES methodology shifts by smoothing differences in employment and earnings between years. Seeding with other years in the same dataset ties a single given year more strongly to the data set as a whole, resulting in fewer outliers that need to be removed or smoothed later in the process.

The graphs below show the effect of multi-year seeding on average and percentile earnings. The lines represent average and percentile earnings for Natural Sciences Managers in Boise, ID for 2012-2017.



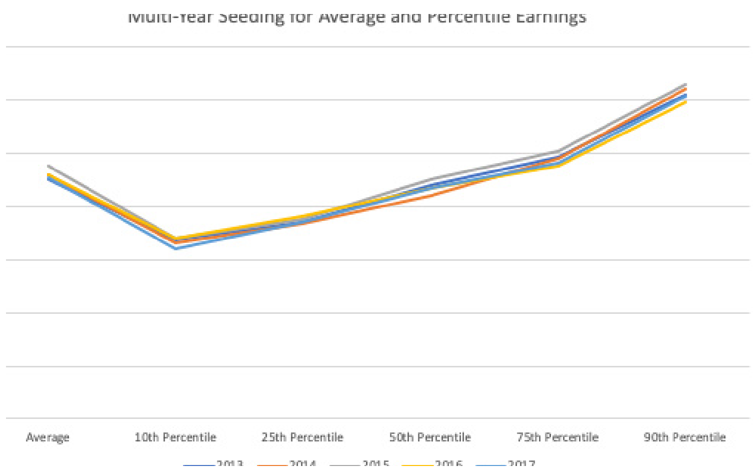
The first graph shows an example in which earnings for a particular occupation in a particular MSA were disclosed in 2013-2016, but were suppressed in 2017. The blue line shows what Emsi's estimates for average and percentile earnings would be under the old methodology, without using the multi-seeding technique where seeds are drawn from other years of data. The years for which the data was disclosed cluster nicely, but the suppressed, single-year-seeded 2017 earnings are fairly different. This is the mysterious OES volatility that many clients notice in Emsi data if they compare across years--individual OES years are quite different from each other.

The next graph shows what happens when multi-year seeding is used to seed the unsuppression for 2017 average earnings--they are brought into line with other years. Additionally, the overall shape and height of the percentile wage curve also benefits from the improved average earnings, since average earnings are used to help unsuppress percentile earnings.



Finally, using multi-year seeding for percentiles as well as for average earnings brings all percentiles into line with the earnings figures for prior years.

Multi-year seeding as an unsuppression tactic helps reduce any volatility introduced into the series as a result of hidden OES methodology changes and contributes to the stability of the OES time series Emsi has built.



Conclusion

The OES time series project allows Emsi to provide its users with more stable, consistent OES employment data over time. It also allows us to offer historical occupation earnings for the first time, and to do so in a way that accounts for the cautions the BLS rightly issues for dealing with multiple years of OES data. Historical occupation earnings have been a frequent request from Emsi clients for many years, and we are excited to be able to extend our offering of relevant, useful, statistically rigorous data to our clients.