# Detailed Overview of Framework

The **location quotient (LQ)** measures the concentration of an industry in a region compared to some benchmark region.

where is the number of establishments in a given industry () and/or region (). The region is the benchmark region overall (entire US). Prior studies indicate that estimated LQs may be unstable, particularly in low-exposure regions.

We can adjust the LQ using the **Buhlmann credibility factor** : . The value of will always be 1, since the we’re examining the ratio of the benchmark region to itself, meaning the adjustment collapses to:

The goal of adjusting the LQ with credibility theory is to “pull” LQ estimates towards the mean by an amount that scales with their level of uncertainty. To estimate the credibility factor (Z), we use the credibility equation:

where is the number of observations in a region (e.g. ); and is a constant reflecting the balance between within-region variability and across-region variability. It is calculated with the **Buhlmann Formula:**

* **EPV** (Expected Value of Process Variance): measures the average variability of industry proportions within individual regions.
  + The proportion of each industry in each region is
  + The variance of these proportions for a given region is
  + The average of these variances across all regions is the EPV:
  + *Intuition:* EPV quantifies how much industry proportions in a region might fluctuate due to randomness/sampling variability. A higher EPV suggests that observed differences within regions are due to randomness and not systematic factors.
* **VHM** (Variance of Hypothetical Means):measures the average variability of industry proportions across all regions.
  + The average proportion of each industry across regions:
  + The variance of these proportions for each industry:
  + The average of these variances across all industries is the VHM:
  + *Intuition:* A higher VHM indicates significant disparities in industry concentrations among regions, reflecting actual regional specialization.

The **parameter K** balances these two components. If K is high (EPV > VHM), random variability within regions dominates, meaning regional data is less credible, and low weights are given to traditional LQs. If K is low (EPV < VHM), systematic differences between regions dominate, meaning regional data is more credible, and more weight is given to regional LQs.

Since and , and the credibility adjustment is , as , , meaning we rely heavily on regional estimates, ignoring the benchmark. As , meaning , we rely heavily on the benchmark, decreasing the weight of regional estimates.

# Observations so Far

Considering both the full US dataset and the Maine dataset, we are seeing most computed values for very close to 1. The average outcome in the Maine data, for example, is 0.99982, implying that VHM > EPV, meaning variability across regions is significantly greater than variability within regions. This suggests differences in industry concentrations across counties compared to the variability within each county, implying regional specialization and a low variability / randomness within individual regions.

Some thoughts:

1. This could be a meaningful result in itself, suggesting that adjusting in this way is not necessary because the data reflects real industry patterns that are well-captured by regional LQs, despite the inherent instability in lower-count regions. Perhaps the use of CIs is a better approach in this case?
2. Alternative credibility methods exist, such as Bayesian credibility models which don’t rely on fixed parameters . Instead, we assume the true is distributed and the regional mean has a prior distribution centered around the benchmark LQ with some variance ). In this setup, the posterior distribution for given the observed data will blend regional LQs with the benchmark LQ, with weights depending on the relative variances and . Regions with a smaller will have posteriors closer to their observed LQs, while those with higher variance will shrink towards the benchmark.