Statistical Methods and Procedures in Estimates of Methane Emissions from Unconventional Wells – A Comment on Bayesian Applications

Jacob Dearmon, Ph.D.
Russell Evans, Ph.D.
## Dataset

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Whole Gas, Average Emissions per Completion (Mcf)</th>
<th>Modified, Average Methane Emissions per Completion (Mcf)</th>
<th>Rounded, Average Methane Emissions per Completion (Mcf)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weatherford</td>
<td>667</td>
<td>555</td>
<td>600</td>
</tr>
<tr>
<td>Industry Data</td>
<td>5,820</td>
<td>4,844</td>
<td>5,000</td>
</tr>
<tr>
<td>Set #1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Devon</td>
<td>11,900</td>
<td>9,905</td>
<td>10,000</td>
</tr>
<tr>
<td>Williams</td>
<td>24,449</td>
<td>20,351</td>
<td>20,000</td>
</tr>
</tbody>
</table>

*Modified emissions are calculated using a methane content value of 0.8324*
Objective

• Develop an interval estimate for the mean methane emission per completion given four summary observations.

• Interval estimation is based on the following assumptions:
  – Data collection design is acceptable
  – Measurement error is not influential
**Interval Estimate:**
**Least Squares Approach**

- **Dependent variable:** Methane Emissions

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-ratio</th>
<th>p-value</th>
<th>Lower 95% C.I.</th>
<th>Upper 95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const.</td>
<td>8900</td>
<td>4168.53</td>
<td>2.1350</td>
<td>0.12243</td>
<td>-4366.13</td>
<td>22,166.1</td>
</tr>
</tbody>
</table>

- **Conclusion:** The result is not statistically significant at the 1%, 5%, or 10% level.
The Bayesian Approach

• **Objective**: Reduce the size of the interval to make the result statistically significant.

• **Method**: Bayesian Econometrics
Bayesian Basics

- $p(\theta|y) \propto p(y|\theta) * p(\theta)$
  - Posterior: $p(\theta|y)$
    - Represents a combination of the data-driven likelihood function and the researchers prior beliefs.
  - Likelihood: $p(y|\theta)$
    - Data-driven density.
  - Prior: $p(\theta)$
    - Researcher’s prior beliefs independent of the data.
• Notice how “tight” the distribution of the prior is relative to the OLS results.
Influence of Prior

\[ \sigma_{EF}^2 = \frac{1}{\frac{1}{\tau^2} + \frac{4}{\sigma^2}} \]

\( \tau \) values for which the null of a mean value of zero is rejected.
Concerns, Considerations and Conclusions

• Variance is assumed to be known
  – Issue: Assumption is not adequately justified. Solution: Bayesian methods can estimate models with an unknown variance in a straightforward manner. A commonly used linear regression model with a natural conjugate prior results in a distribution that would follow a t-distribution with a wider and more representative interval once the variance has been integrated out.

• Small dataset
  – Issue: Only a few data points which suffer from measurement error. Solution: Collect more data at the well level using a survey/process designed explicitly to measure vented methane emissions.

• Prior
  – Issue: Prior, rather than the data, is driving the results. Solution: Combine a diffuse or uninformative prior with the data to allow for empirically grounded conclusions.