

## Quick-scoring guide for the Maryland Scientific Method Scale

| Method   | Maximum SMS score<br>(method, implementation)  | Adjusted SMS score<br>(method, implementation)   |  |
|--|--|--|--|
| <b>Randomised Control Trial (RCT)</b><br><i>a.k.a.</i><br>Field Experiment           | 5, 5 if <ul style="list-style-type: none"> <li>Randomisation is successful</li> <li>Attrition carefully addressed or not an issue</li> <li>Contamination not an issue</li> </ul>   | 5, 4 if <ul style="list-style-type: none"> <li>One of the criteria is severely violated</li> </ul> 5,3 if <ul style="list-style-type: none"> <li>Two or more of the criteria are severely violated</li> </ul>                                      |  |
| <b>Instrumental Variable (IV)</b><br><i>a.k.a.</i><br>Two-Stage Least Squares (2SLS) | 4, 4 if instrument <ul style="list-style-type: none"> <li>Relevant (explains treatment)</li> <li>Exogenous (not explained by outcome)</li> <li>Excludable (does not directly affect outcome)</li> </ul>  | Scored as per underlying method if <ul style="list-style-type: none"> <li>Instrument invalid</li> <li>e.g cross section with invalid IV scores 2; difference-in-difference with invalid IV scores 3 (see below)</li> </ul>                         |  |
| <b>Regression Discontinuity Design (RDD)</b>   | 4, 4 if <ul style="list-style-type: none"> <li>Discontinuity in treatment is sharp (e.g. strict eligibility requirement) or fuzzy discontinuity method used</li> <li>Only treatment changes at boundary</li> <li>Behaviour is not manipulated to make the cut-off</li> </ul> | Scored as per underlying method if <ul style="list-style-type: none"> <li>Discontinuity conditions severely violated</li> <li>e.g cross section with invalid IV scores 2; difference-in-difference with invalid IV scores 3 (see below)</li> </ul> |  |
| <b>Difference in differences (DID)</b><br><i>a.k.a.</i><br>Diff in diff              | 3,3 if <ul style="list-style-type: none"> <li>Control group would have followed same trend and treatment group</li> <li>Known time period for treatment</li> </ul>   | 3, 2 if <ul style="list-style-type: none"> <li>Either of the criteria is not satisfied</li> </ul>  |  |
| <b>Panel methods</b>   | <b>Panel Fixed Effects (FE)</b>  | 3, 3 if <ul style="list-style-type: none"> <li>Fixed effect is at the unit of observation</li> <li>Year effects are included</li> <li>Appropriate time-varying controls are used</li> </ul>  | 3, 2 if <ul style="list-style-type: none"> <li>One or more of the three criteria is not satisfied</li> </ul> |
|  | <b>First Differences (FD)</b>  | 3, 3 if <ul style="list-style-type: none"> <li>Year effects are included</li> <li>Appropriate time-varying controls are used</li> </ul>  | 3, 2 if <ul style="list-style-type: none"> <li>Either of the two criteria is not satisfied</li> </ul>        |
|  | <b>Arellano-Bond (AB)</b>  | 3, 3 if<br>Year effects are included <ul style="list-style-type: none"> <li>Appropriate time-varying controls are used</li> </ul>  | 3, 2 if <ul style="list-style-type: none"> <li>One of the two criteria is not satisfied</li> </ul>           |

| Method   |   | Maximum SMS score<br>(method, implementation)  | Adjusted SMS score<br>(method, implementation)  |
|--|---|--|---|
| <b>Hazard Regressions</b>  | <b>Mixed Proportional Hazards (MPH)</b> | 4, 4 if <ul style="list-style-type: none"> <li>Key assumption of 'no anticipation' holds</li> <li>Variation in timing (e.g. people start training at different times relative to becoming unemployed)</li> </ul> | 4, 3 if <ul style="list-style-type: none"> <li>Anticipation of treatment likely</li> <li>Little variation in timing</li> </ul> 4, 2 if <ul style="list-style-type: none"> <li>Neither of the criteria is satisfied</li> </ul> |
|  | <b>Proportional Hazards (PH)</b>        | 3,3 if <ul style="list-style-type: none"> <li>Adequate control group is established</li> <li>Treatment date is known and singular</li> </ul>   | 3,2 if <ul style="list-style-type: none"> <li>One of the two criteria is not satisfied</li> </ul>   |
| <b>Heckman Two-Stage Approach (H2S)</b><br>Or<br><b>Control Function (CF)</b>        | <b>With IV</b>                          | 4,4 if <ul style="list-style-type: none"> <li>Selection equation includes an IV that satisfies the three criteria (see IV)</li> </ul>  | Scored as per underlying method if <ul style="list-style-type: none"> <li>Instrument invalid</li> <li>e.g cross section with invalid IV scores 2; difference-in-difference with invalid IV scores 3 (see below)</li> </ul>    |
|  | <b>With DID or panel method</b>         | 3,3 if <ul style="list-style-type: none"> <li>Selection equation includes relevant observable variables</li> <li>DID or panel criteria met</li> </ul>  | 3, 2 if <ul style="list-style-type: none"> <li>One of the criteria is not satisfied</li> </ul>  |
|  | <b>Cross-sectional</b>                  | 2, 2 if <ul style="list-style-type: none"> <li>Selection equation includes relevant observable variables</li> </ul>  | 2, 1 if <ul style="list-style-type: none"> <li>Selection equation does not include relevant observable variables</li> </ul>   |
| <b>Propensity Score Matching (PSM)</b><br><i>a.k.a.</i><br>Matching                  | <b>With DID or panel method</b>         | 3, 3 if <ul style="list-style-type: none"> <li>Matching criteria satisfied (see below)</li> <li>DID or panel criteria satisfied</li> </ul>   | 3, 2 if <ul style="list-style-type: none"> <li>One of the criteria is violated</li> </ul>   |
|  | <b>Cross-sectional</b>                  | 2, 2 if <ul style="list-style-type: none"> <li>Good matching variables (i.e. relevant to selection)</li> <li>Significant common support</li> </ul>   | 2, 1 if <ul style="list-style-type: none"> <li>One of the criteria is violated</li> </ul>   |
| <b>Cross-sectional regression</b>  |   | 2, 2 if <ul style="list-style-type: none"> <li>Adequate control variables are used</li> </ul>  | 2, 1 if <ul style="list-style-type: none"> <li>Inadequate control variables</li> </ul>  |
| <b>Before-and-after</b>  |   | 2, 2 if <ul style="list-style-type: none"> <li>Adequate control variables are used</li> </ul>  | 2, 1 if <ul style="list-style-type: none"> <li>Inadequate control variables</li> </ul>  |
| <b>Stated effects/impact</b><br><i>a.k.a.</i><br>Additionality<br>Stated experiences |   | Not SMS scoreable  |   |
| <b>Impact modelling</b>  |   | Not SMS scoreable  |   |