数据包络分析（DEA）常见问题总结

1、问：DEA投入和产出指标需不需要标准化处理？  
不需要进行标准化处理。如果要进行标准化处理，要注意数据变换对分析结果的影响。  
1）指标的所有数据都除以指标的最大值：这种变换与改变数据测量单位性质相同，绝大多数DEA模型与数据的测量单位无关(unit-invariant),因此这种变换之后，DEA分析结果保持不变。  
2）离差标准化，0-1标准化(0-1 normalization)：是对原始数据的线性变换，使变换后的数据在[0,1]区间。数据变换之后，DEA分析结果通常会发生改变。只有个别DEA模型在数据进行线性变换之后结果保持不变或部分结果（例如DMU是否有序的定性结果）保持不变（unit-invariant and translationinvariant）。  
  
2、问：DEA模型中VRS和CRS是什么意思？  
VRS表示规模报酬可变，CRS表示规模报酬不变。  
  
3、问：DEA模型中VRS产出导向模型和投入导向模型的区别？  
在VRS产出导向模型中，如果一个DMU的某项投入小于任何其他DMU，则该DMU效率值必定等于1（西藏就是这种情况）；在VRS投入导向模型中，如果一个DMU的某项产出大于任何其他DMU，则该DMU效率值必定等于1。由于没有可比标杆，因此只能默认为有效。并且在以上模型的超效率模型中，该DMU无可行解。  
  
4、问：DEA模型对投入产出指标和DMU（决策单元个数）的要求是什么？  
一般而言，为了保证DEA对DMU效率的区分能力，DMU的数量应不少于投入和产出指标数量的乘积，同时不少于投入和产出指标数量的3倍。  
  
5、问：DEA模型存不存在共线性问题？  
共线性问题只有参数方法（例如回归）才涉及，DEA为非参数方法，和共线性问题不沾边。  
  
6、问：DEA模型技术效率到底是什么意思？  
E就是Efficiency, 再准确些说就是Technical Efficiency，中文是“技术效率”，其含义为被评价DMU的生产达到当前（该DMU数据所处的时期）技术水平的程度。再具体些说，就是在投入一定的条件下，被评价DMU的产出达到最大化的程度，通常表示为被评价DMU的产出与在当前（该DMU数据所处的时期）技术水平下所能达到的最大产出的比值；或者在产出一定的条件下，被评价DMU的投入达到最小化的程度，通常表示为在当前（该DMU数据所处的时期）技术水平下所能实现的最小投入与被评价DMU的投入的比值。以上解释中，有一个关键点，就是技术效率的测量总是与当前（该DMU数据所处的时期）技术水平绑定的，从时间上讲，在评价某一DMU的技术效率时，用于参比的“技术”与被评价DMU的数据是来自同一时期的。也就是说，技术效率的测量，不存在跨期的比较。技术效率的测量就是被评价DMU与同一时期内"表现最好"的DMUs（也就是构成前沿的DMUs)进行比较。技术是一个抽象的概念，在DEA分析中，技术具体化为“前沿”。可以这样理解，技术就是前沿，前沿就代表技术（更准确地说，在DEA中，技术是用前沿所包裹的生产可能集来表示的，在此为方便表述，将技术简化为前沿）。进一步说，技术效率的测量就是被评价DMU与“同一时期内”的一组“表现最优”的DMUs进行比较。  
  
  
7、问：DEA模型中的技术效率变化是什么意思呢？  
EC是Efficiency Change，是技术效率值在两个时期发生的变化。某一被评价DMU的EC，就是该DMU在时期t+1的技术效率值和时期t的技术效率值之间的比值。该DMU在t+1期的技术效率值由该DMU用t+1期的投入产出数据与t+1期内表现最优的DMUs进行比较得出，该DMU在时期t的技术效率值由该DMU用t期的投入产出数据与t期内表现最优的DMUs进行比较得出。总之，技术效率的测量不跨期。用面板数据分析得出的EC，同样可以用以下等价的方法计算得出：1）将包含T期的面板数据每个时期分离成一个单独的截面数据，共T个单独的数据表，然后分别单独分析这T个截面数据，每个DMU都会得出T个效率值，计算每个DMU相邻两期效率值的比值，这就是每个DMU的EC。  
  
8、问：DEA模型中的技术变化是什么意思?TC又是如何测量的呢？  
TC，Technological Change，是指两个时期的技术变化。前面已经说过，技术在DEA中具体化为前沿，那么技术变化就具体化为两个前沿发生的变化。前沿可以往前移动（技术进步），也可以往后移动（技术退步）。单个时期（截面数据）对应的是一个技术（一个前沿），不存在技术变化的问题，技术变化必然是涉及到要进行比较的两个技术（也就是两个时期的前沿），因此对TC的分析必然是跨期的。将包含T期的面板数据每个时期分离成一个个单独的截面数据，对每个时期的截面数据进行单独分析是不可能得出TC的。  
接下来的问题就是如何测量TC，TC是两个时期前沿发生的变化，前沿是由多个DMUs所构成的曲线（面）。在DEA分析中，前沿的变化是通过一个被评价DMU分别与两个前沿做比较来间接反映出两个前沿之间的相对变化：该DMU与一个前沿进行比较，得出的分数越低，说明该DMU与此前沿之间的距离越远。这样通过比较两个前沿与一个DMU之间的距离就能得出两个前沿之间所发生的相对变化（这个TC值可以理解为是从这个DMU的视角出发所看到的两个前沿之间的距离）。每个DMU的“视角”是有所差别的（用图示表示的更清楚），在数据中有多少个DMU，就能得出多少个TC，每个DMU得出的TC都是从该DMU所处的位置去反映两个前沿之间所发生的变化。这就是为什么两个时期的前沿进行比较，得出的TC不是一个固定数值，而是有N个数值。  
  
9、问：经济学中的影子价格和DEA中的影子价格的区别？  
以投入的影子价格为例，在经济学中计算影子价格时，目标函数通常是产出（真实产出），此时得出的影子价格含义就是投入的边际产出（每增加一个单位投入，目标函数即产出的增量）；而在DEA模型中（投入导向模型），目标函数是效率值，同时也是虚拟产出值（即各产出加权之和），并非真实产出值，因此影子价格的含义对应为每增加一个单位投入，效率值的增量（负值），或虚拟产出的增量（负值）。这个虚拟产出是DEA模型中所特有的。二者计算原理相同，但实际含义差别明显。讨论指标时，一定要注意指标所在的“环境”，差之毫厘，失之千里。当只有单投入单产出时，这两个影子价格之间似乎可以互相转换计算。  
DEA结果中的dual price（shadow price）是与模型的目标函数关联的，其含义为：投入（产出）每增加一个单位，目标函数值的变化值。所以要使用DEA包络模型结果中的对偶价格（对应对偶乘数模型结果中的投入产出权重），首先得明确所使用的DEA模型的目标函数。否则，不能轻易使用这些结果。  
  
10、问：DEA的投入或产出指标多增加一个指标，对结果的影响？  
在原模型基础上，增加一个或多个指标（无论是投入还是产出，无论是好产出还是坏产出）后，效率值一定是增大或不变，不可能减小！在此说的是传统的径向模型（效率值增大或不变）和方向距离函数模型（Beta值减小或不变）。这是可以证明的。其他距离类型，例如SBM，则结果可能增加也可能减小。