如何用简单的例子解释什么是 Generalized Method of Moments (GMM)?

GMM estimator 在asset pricing领域十分有用，能不能用简单的例子解释下？最近在学习Cochrane的asset pricing…显示全部

11 个回答 默认排序

慧航
2015 新知答主
关注

既然被邀请和提到，在这里我来写一个最简单的GMM快速入门手册吧，因为这个技术听起来非常的高大上，但其实非常简单。如果你有本科的统计知识，看懂下文是不成问题的。

GMM的全名是Generalized Method of Moments，也就是广义矩估计。只看这个名字的话，如果去掉「广义」这个词，可能学过本科统计的人都认识，就是「矩估计」。

矩估计是什么呢？简单的说，就是用样本矩代替总体矩进行统计推断的方法。一个最基础的例子是正态总体的参数估计问题。如果$x_i \sim N(\mu,\sigma^2)$，如何估计μ和σ呢？

本科的统计学一般会介绍两种方法：极大似然估计和矩估计。其中矩估计是我们今天的主角。观察到：

$$Ex_i=\mu,\ E(x_i^2)=\mu^2+\sigma^2$$

而根据大数定理，在一定的条件下，我们有：

$$\bar{x_i}-\mu=o_p(1),\bar{x_i^2}=\mu^2+\sigma^2+o_p(1)$$

也就是说，当样本量足够大的时候，样本矩与总体矩只差了一个无穷小量，那么我们是不是可以用样本矩代替总体矩得到参数的估计呢?

按照上面的思路，我们把op(1)去掉，同时把未知的总体参数写成其估计值，也就是带hat的形式，我们得到了：

$$\hat{\mu}=\bar{x_i},\hat{\sigma^2}=\bar{x_i^2}-\bar{x_i}^2$$

如此，我们得到了两个总体矩的点估计。在这个简单的例子里面，你只要把上面的大数定理的结论带到上面两个式子里面，很容易的就可以证明出两个点估计是一致的估计量。

当然，值得注意的是，即便我使用的是矩条件，σ的估计也不是无偏的。一般而言，除了特殊情况，不管是MLE还是MM还是GMM，都不一定可以得到无偏的估计量。特别是在比较复杂的应用里面，一致就很不错了，无偏性的讨论真的繁琐。

好了，上面是矩估计，非常简单是吧？但是什么又是广义矩估计呢？

在上面的例子中，我们只使用了两个矩条件。然而我们知道，正态分布的矩是有无穷多个可以用的，那么我们是不是可以使用更多的矩条件呢？

但是有个问题不好解决。在这个例子里面，我们有两个未知参数，如果只使用一阶矩，那么只有一个方程解两个未知数，显然是不可能的。像上面一样，我们用两个矩条件解两个未知数，就解出来了。然而，当我们用一到三阶矩，总共三个方程求解的时候，三个方程求解两个未知数，可能无解。

方程数多了，反而没有解了，为什么呢？其实很简单，用三个方程中的任意两个方程，都可以求出一组解，那么三个方程我们就可以求出三组解。所以应该如何把这些矩条件都用上呢？

到这里我们不妨引入一些记号。还是使用上面的例子，我们把上面的三个矩条件写到一个向量里面去，记：

相关问题
如何理解“generalized empirical method”？ 0 个回答
研究derived scheme这样听起来很高深的概念有什么用吗? 0 个回答
apologize创作背景? 0 个回答
Apologize 和 apologise 哪个是英式哪个是美式? 2 个回答

$g(x_i,\theta)=\left[x_i-\mu,x_i^2-\mu^2-\sigma^2,x_i^3-\mu^3-3\mu\sigma^2\right]',\theta=\{\mu,\sigma^2\}$

我们可以得到一个3*1的列向量，并且：

$Eg(x_i,\theta)=0$

上面就是我们要用的矩条件。而根据上面的思路，用其样本矩代替总体矩：

$\frac{1}{N}\sum_i g(x_i,\hat{\theta})=0$

解这个方程应该就可以得到参数θ的估计。但是正如上面所说的，三个方程两个未知数，并不能确保这个方程有解，所以必须想一些其他办法。

一个比较自然的想法是，上面的矩条件等于0，虽然我不太可能保证三个方程同时等于0，但是仿照OLS，我们可以让他们的平方和最小，也就是：

$\min_{\hat{\theta}}\left[\frac{1}{N}\sum_i g(x_i,\hat{\theta})\right]'\left[\frac{1}{N}\sum_i g(x_i,\hat{\theta})\right]$

这样我们就能保证三个矩条件的样本矩都足够贴近于0，当然不可能同时为0。这样不就综合使用了三个矩条件的信息么？

更一般的，由于上面的g函数是一个3*1的列向量，我们可以使用一个权重矩阵W来赋予每个矩条件以不同的权重：

$\min_{\hat{\theta}}\left[\frac{1}{N}\sum_i g(x_i,\hat{\theta})\right]'W\left[\frac{1}{N}\sum_i g(x_i,\hat{\theta})\right]$

只要这个W是一个正定矩阵，那么仍然可以保证每个样本矩都足够贴近于0。

那么问题来了，既然对W的要求只要求正定矩阵，那么使用不同的权重矩阵就有可能得到不同的结果。问题是，有没有一个最优的权重矩阵呢？当然是有的。可以证明，最优的权重矩阵应该是：

$[Eg(x_i,\theta)g(x_i,\theta)']^{-1}$

使用这个权重矩阵，就得到了最有效的估计。

比如上面的例子，用gretl分别估计两个矩条件、三个矩条件使用单位阵作为W、三个矩条件使用最优权重矩阵做估计：

```
nulldata 1000
set seed 1988
series x=randgen(N,1,2)
series x2=x^2
series x3=x^3
series e
series e2
series e3
scalar mu=0
scalar sigma2=1
matrix W2=I(2)
gmm
    series e=x-mu
    series e2=x2-sigma2-mu^2
    orthog e; const
    orthog e2; const
    weights W2
    params mu sigma2
end gmm
matrix W3=I(3)
scalar mu=0
scalar sigma2=1
gmm
    series e=x-mu
    series e2=x2-sigma2-mu^2
    series e3=x3-3*mu*sigma2-mu^3
    orthog e; const
    orthog e2; const
    orthog e3; const
    weights W3
    params mu sigma2
end gmm
```

```
scalar mu=0
scalar sigma2=1
gmm
    series e=x-mu
    series e2=x2-sigma2-mu^2
    series e3=x3-3*mu*sigma2-mu^3
    orthog e; const
    orthog e2; const
    orthog e3; const
    weights W3
    params mu sigma2
end gmm --iterate
```

首先是使用两个矩条件的结果：

```
Model 1: 1-step GMM, using observations 1-1000

             estimate   std. error     z      p-value
  -----------------------------------------------------
  mu         1.11563    0.0633519    17.61   2.06e-69  ***
  sigma2     4.01346    0.176158     22.78   6.72e-115 ***

  GMM criterion: Q = 1.27242e-29 (TQ = 1.27242e-26)
```

为什么两个矩条件的时候不使用最优权重矩阵呢？因为两个未知参数，两个矩条件，不存在过度识别的问题，存在唯一解的，所以不管使用任何的正定矩阵，得到的结果都是一样的。

三个矩条件，这个时候使用什么样的权重矩阵就不一样了。先使用单位阵作为权重矩阵：

```
Model 2: 1-step GMM, using observations 1-1000

             estimate   std. error     z     p-value
  ----------------------------------------------------
  mu         1.09170    0.0844553    12.93   3.20e-38 ***
  sigma2     4.05716    0.209200     19.39   8.72e-84 ***

  GMM criterion: Q = 0.000663476 (TQ = 0.663476)
```

这里需要注意的是，即使使用了更多的矩条件，估计量的standard error还是变大了。感兴趣的可以做一个蒙特卡洛模拟试试，一定是会变大的。为什么呢？因为没有使用最优的权重矩阵，所以使用单位阵作为权重矩阵得到的结果不是最有效的。那么如果使用最优的权重矩阵呢？结果：

```
Model 3: Iterated GMM, using observations 1-1000

             estimate   std. error     z      p-value
  -----------------------------------------------------
  mu         1.11486    0.0633224    17.61   2.21e-69  ***
  sigma2     4.00920    0.175836     22.80   4.50e-115 ***

  GMM criterion: Q = 0.000160784 (TQ = 0.160784)
  J test: Chi-square(1) = 0.160784 [0.6884]
```

嘿！standard error是变小了，但是跟使用两个矩条件的好像没有什么本质变化啊？为什么呢？

因为这里举的这个例子太特殊了，我们使用的前两个矩条件，刚好是一个充分统计量，也就是说，使用额外的矩条件不会带来附加信息的。但是如果是其他情况，一般来说更多的矩条件是可以带来更多的信息的，比如工具变量的回归。

另外如果细心观察，最后一张表格多了一个J-test。这又是啥呢？

这个东西就比较有意思了。知道现在，我们都是假设使用的矩条件成立，那么这些矩条件真的是成立的么？未必啊。比如，如果x本来就不服从正态分布，那么使用上面的估计显然是错的。那么是不是可以检验矩条件是否成立呢？

一般来说，如果你有K个未知的参数，以及K个矩条件，那么矩条件是不能检验的。但是如果你有更多的矩条件，那么就有了检验的可能。这个检验的直觉很简单，比如上面的例子里面，我们有3个矩条件。我可不可以先使用前两个矩条件估计这两个参数，然后把这两个参数带入到第三个矩条件里面，看看是不是充分接近于0，如果充分接近，那么看来这三个矩条件彼此印证了。

实际使用的时候没有那么麻烦。可以证明，当使用了最优的权重矩阵的时候，GMM的目标函数渐进服从卡方分布，因而只要检验这个卡方分布就可以了，也就是上面的J-test。p-value为0.6884，看来这三个矩条件没有矛盾的地方。

但是一定要注意，即使通过了这个检验，也不代表矩条件一定是成立的，因为有可能三个矩条件都是错的，只不过错的方向是一致的

的，但第四阶就不一样了。因而通过这个检验不代表x一定服从正态分布。当然，如果通不过，可以比较自信的说，x不服从正态分布。

比如，我们把上面的数据生成过程改为gamma分布，得到的结果：

```
Model 3: Iterated GMM, using observations 1-1000

             estimate   std. error     z      p-value
  ----------------------------------------------------
  mu         1.83716    0.0538519    34.12   4.40e-255 ***
  sigma2     2.43215    0.142548     17.06   2.85e-65  ***

  GMM criterion: Q = 0.0280774 (TQ = 28.0774)
  J test: Chi-square(1) = 28.0774 [0.0000]
```

p-value为0.0000，拒绝了原假设，也就是说，三个矩条件不同时成立，数据很有可能不是从正态分布中生成的。

计量经济学的很多很多问题基本都可以归结为GMM的问题。从最简单的OLS、2SLS到稍微复杂一点的面板数据、动态面板等等，本质上都是在找矩条件。比如工具变量的2SLS，可以发现矩条件不过就是：

$$E[(y_i - x_i'\beta)z_i] = 0$$

套一下上面的公式，最优权重矩阵(的逆)为：

$$E[(y_i - x_i'\beta_0)z_i z_i'(y_i - x_i'\beta_0)'] = E[e_i^2 z_i z_i'] = \sigma^2 E z_i z_i'$$

带入到目标函数中，就得到了2SLS。

甚至，一些其他的估计量，比如MLE、M-estimator等，在一定的条件下也可以转化为GMM，因为这些估计量的一阶条件可以看成是矩条件。所以GMM也就变成了一个统一的框架。

为什么GMM这么受欢迎呢？因为GMM把复杂的统计过程抽象化成为一个（看似）简单的过程：找矩条件。只要你能找到矩条件，你就能估计。GMM把估计的繁琐细节全都抽象了，面对一个模型，你所需要做的所有事情就是找到矩条件，证明这个模型是可以识别的，然后什么也不用管，一股脑儿塞进去，结果就出来了。

所以呢如果你去看一些稍微复杂的模型，基本都可以归结为矩条件。

至于题主提到的资产定价，刚好Gretl提供了一个可以使用的数据集和code。资产定价最简单的模型应该就是C-CAPM了，其重要结论就可以直接归结为这么一个矩条件：

$$E\left[\delta \frac{r_{j,t+1}}{p_{j,t}}\left(\frac{C_{t+1}}{C_t}\right)^{\alpha-1} \middle| \mathcal{F}_t\right] = 1,$$

其中Ft为第t期所知道的所有信息，包括Ct、rt等等。所以根据这个式子，如果令

$$e_t = \delta \frac{r_{j,t+1}}{p_{j,t}}\left(\frac{C_{t+1}}{C_t}\right)^{\alpha-1} - 1$$

那么e_t跟Ct、rt等等都是正交的，自然可以作为矩条件来用。

Gretl自带了Hall的数据集，在user guide第206页开始给出了说明和代码，以及结果，感兴趣的可以去看看，很简单的一个程序。

我猜想上面的两个例子已经足够简单了，特别是正态分布的例子，应该不可能更简单了哈哈～

Huang Zibin
面对现实，忠于理想

GMM简直是计量的良心

它可以涵盖几乎所有常用的estimator

OLS, IV, 2SLS, GLS, RE, FE, SU