构建反事实的面板方法选择性综述

2025 年 6 月 14 日，由北京大学中国经济研究中心（CCER）、北京大学国家发展研究院主办的第九届 CCER 夏季国际研讨会在国家发展研究院（承泽园校区）进行。南加州大学教授萧政（Cheng Hsiao）应邀出席并做主题演讲。本文根据演讲内容整理。

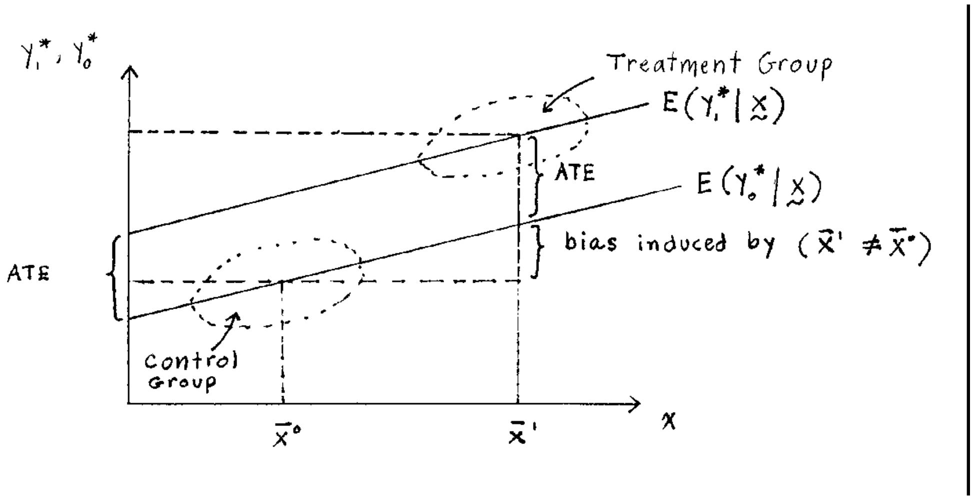
我非常荣幸能有机会在 CCER 展示我的工作，感谢沈艳教授的精彩介绍。正如沈艳教授所说，我长期从事面板数据的研究，面板数据相比传统的横截面数据和时间序列数据具有诸多优势。首先，我将介绍面板数据并重点阐述其在实证分析中的优越性。随后，我将讨论测量处理效应时需要关注的核心问题，并进一步分析传统经济学或计量经济模型在利用面板数据识别经济模型参数以测量处理效应时的相关方法。接下来我想讨论，假设我们只关注预测，而不是识别经济模型中的参数，那么，利用面板数据进行无偏预测有哪些简单方法？我将主要以单个受处理个体为例来介绍相关的计量方法，并讨论如果存在多个个体接受处理会涉及的问题。最后，我想就此做出一些总结性发言。

使用面板数据具有诸多优势

我们假设在观测到的面板中，任一时点存在N个截面个体，每个个体在T个时间段上被观测。与传统的截面数据或时间序列数据模型相比，我认为面板数据具有多方面的优势。

第一个优势，众所周知，在回归分析中，我们将结果表示为某些协变量的函数与随机误差项之和。其中一个标准假设是，误差项代表被遗漏因素的影响，与经济学家认为影响结果的重要可观测因素不相关，但这一假设相当严格，这些被遗漏的因素很有可能与协变量相关。面板数据可以处理这种相关性。

第二个优势是在测量处理效应时关注的核心问题。下图中，横轴上是认为会影响结果的变量因素x，纵轴上，使用 y_1^*表示接受处理的个体的结果，y_0^*表示未接受处理的个体的结果。假设上方的曲线表示在不同 x 值下接受处理的个体的结果，下方的曲线则表示未接受处理的个体的结果。那么给定可观测因素 x，处理效应就是上方曲线与下方曲线在 x 处的差值。这就是衡量处理效应的方法。



但是，如果你比较接受处理的个体与未接受处理的个体，会发现两者在协变量x上可能存在不同。因此，二者之间的差异由两部分组成：一是平均处理效应，二是由于协变量x差异所导致的偏差。这种差异可能源于两个方面：一方面是条件协变量的不同，另一方面则是处理组和控制组的个体并非随机分配。如何控制这些差异，是测量处理效应时的核心问题。在这种情况下，面板数据为解决因为协变量x的差异带来的问题提供了更好的机会，这类问题被称为"基于可观测变量的选择"或因处理组和控制组非随机分配，Heckman 也称其为样本选择问题，相关文献十分丰富。面板数据为解决上述样本选择问题提供了更广阔的视角。

在评估政策，特别是宏观政策效果时，传统方法通常将重点放在总体效应上。然而，总体效应的加总在典型的经济理论中，通常依赖于"代表性个体"假设，即认为每个人都是类似的。但这一假设能否成立，实际上取决于个体之间是否真正同质。通过面板数据的分析，可以更好地区分同质性与异质性的问题。

此外，与横截面数据相比，面板数据具有独特的优势。利用横截面数据来测量处理效应时，通常只能判断是否存在处理效应。但如果考虑到人类行为的惯性，以及制度和技术变化所带来的调整僵化，处理效应可能是随时间逐步演化的，给定某一时点进行测量未必能够体现。面板数据由于包含了时间序列信息，可以捕捉个体内部的动态变化，这对于研究政策变动对结果的演变过程具有明显优势。

最后，仅从预测的角度来看，我们可以测量处理效应，但这种方法没有经济解释。这也是我认为人工智能或机器学习方法与经济模型之间的根本区别之一。经济模型试图解释我们观察到的现象发生的原因，以及如果外部条件变化，人们将如何做出反应。在使用面板数据时，即使采取纯粹的预测方法来衡量处理效应，由于个体之间的差异，也有可能发现解释不同个体对政策变化反应的根本因素。我个人认为这正是面板数据的优势所在——它能够结合个体差异和个体内部的动态变化，为这些问题提供新的洞见。

处理效应的测量

在我进行综述之前，我想先补充一些记号设定。假设有N个截面单位，每个单位在T个时间段内被观测。在这N个截面单位中，假设有N_1个单位在最初的几个时期未接受处理，而在此之后接受处理。对于这些个体，我们可以观测到其在处理前后的结果，我将这些个体称为处理组个体。其余$N - N_1$个未接受处理的单位，被称为控制组个体。

首先，处理效应的测量只是一个预测问题。

假设用y_{it}^1表示第 i个个体在第t时期接受处理时的结果，用y_{it}^0表示第i个个体在第t时期未接受处理时的结果，那么处理效应Δ_{it}即为 $y_{it}^1 - y_{it}^0$。然而，对于在任意时点的任意个体，观测数据只会记录其接受处理或未接受处理时的结果，而不会同时获得该个体在两种情况（即同时接受和未接受处理）下的结果。

在这种情况下，使用虚拟变量$d_{it} = 1$表示第i个个体在第t期接受处理，$d_{it} = 0$表示第i个个体在第t期未接受处理，那么观察到的数据为$y_{it} = d_{it}y_{it}^1 + (1 - d_{it})y_{it}^0$。你无法同时观测到$y_{it}^1$和$y_{it}^0$。因此，在衡量第$i$个个体在第$t$期的处理效应时，你并没有同时拥有这两个数值。假设在这个时期，你观测到了y_{it}^1，而没有y_{it}^0，此时你需要预测y_{it}^0的值，从而估计处理效应。

估计的处理效应的期望值就是y_{it}^1的期望值减去y_{it}^0的预测值的期望。如果将y_{it}^0的预测值进行分解，可表示为个体在未接受处理时的期望值与预测误差之和。因此，处理效应的估计方法本质上是在于如何让个体未接受处理时的预测值尽可能接近其实际未接受处理时的结果。

$$E(\hat{\Delta}_{it}|y_{it}^1) = E(y_{it}^1 - \hat{y}_{it}^0) = E[(y_{it}^1 - y_{it}^0) + (y_{it}^1 - \hat{y}_{it}^0)] = E(\Delta_{it}|y_{it}^1) + E(y_{it}^0 - \hat{y}_{it}^0), \qquad (2.4)$$

and

$$\mathrm{Var}(\hat{\Delta}_{it}|y_{it}^1) = E\left[(y_{it}^0 - \hat{y}_{it}^0)^2\right]. \qquad (2.5)$$

In other words, the bias and variance of $\hat{\Delta}_{it}$ conditional on y_{it}^1 only depend on the bias and error variance of the prediction error of $\hat{y}_{it}^0$ (or $\hat{y}_{it}^1$ if y_{i0}^1 is observed). That is, to obtain accurate measurement of Δ_{it} is fundamentally an issue of getting good prediction of y_{it}^0. However, there is no realized y_{it}^0 to evaluate how close $\hat{y}_{it}^0$ is to y_{it}^0. Any $\hat{y}_{it}^0$ is a *counterfactual*. There is no

然而，预测和构建反事实之间是有区别的。例如，你预测中国经济在 2025 年底的增长率为 5%。如果实际增长率为 5.2%，你可以将预测值与实际结果进行比较。而在处理效应的测量构建中，y_{it}^0从未被实现。你无法判断你的预测值是否准确。这也是我们称之为反事实的原因，因为没有额外的数据可以验证你的方法是否准确。在这种情况下，我认为，选择特定反事实的估计方法时，其基本假设要与观测到的y_{it}^0的数据生成过程相兼容。给定数据生成过程和估计方法的基本假设，我们可以评估估计方法是否合适，但至于数据生成过程是否真的与观测数据兼容，那又是另外一个问题。

因果方法

我回顾基于单个受处理个体的基本方法，以突出试图识别影响原因的传统经济学模型与我称之为纯预测方法的区别。我假设只有一个个体接受了处理，而其余个体均未接受处理。在第

1 期到第T期期间，没有任何个体接受处理。在第$T+1$期，第一个个体接受处理。此时对于第一个个体，我只能观察其接受处理后的结果。对于其他未接受处理的个体，也是如此。假设第i个个体在第t期接受处理后的结果y_{it}^1可以表示为$g_1(x_{it})$加上一个随机误差项ϵ_{it}^1，该误差项反映了所有被忽略因素对结果的影响。类似地，y_{it}^0可以表示为$g_0(x_{it})$加上ϵ_{it}^0。g_1和g_0均源自某些经济学模型。标准假设认为，被遗漏因素的影响与可观测因素不相关。那么条件于x时的平均处理效应就可以被简单表达为$g_1(x)-g_0(x)$。

the sum of some observed covariates $\boldsymbol{x}_{it}$ and the impact of unobserved factors represented by the error terms

$$y_{it}^1 = g_1(\boldsymbol{x}_{it}) + \varepsilon_{it}^1, \tag{3.1}$$

$$y_{it}^0 = g_0(\boldsymbol{x}_{it}) + \varepsilon_{it}^0, \tag{3.2}$$

where the error terms, ε_{it}^1 and ε_{it}^0 are typically assumed to be uncorrelated with $\boldsymbol{x}_{it}$,

$$E(\varepsilon_{it}^1|\boldsymbol{x}_{it}) = E(\varepsilon_{it}^1) = 0 \tag{3.3}$$

and

$$E(\varepsilon_{it}^0|\boldsymbol{x}_{it}) = E(\varepsilon_{it}^0) = 0. \tag{3.4}$$

Then the average treatment effects conditional on $\boldsymbol{x}$, ATE $(\boldsymbol{x})$, is just $E(y_{it}^1) - E(y_{it}^0)$,

$$\text{ATE}(x) = g_1(\boldsymbol{x}) - g_0(\boldsymbol{x}). \tag{3.5}$$

然而，个体是否接受处理通常由传统的二元选择模型来描述。观测到的d_{it}为 1 还是 0 取决于其潜在响应函数$d_{it}^* = h(z_{it}) + v_{it}$大于零或小于零，其中$d_{it}^*$可表示为某些条件协变量的函数加上误差项。这些随机误差项与条件协变量不相关，且其均值为零。

However, since the observed data take the form (2.2). Suppose, the treatment status dummy or participation decision model for d_{it} can be postulated by introducing a latent response function,

$$d_{it}^* = h(\boldsymbol{z}_{it}) + v_{it},\ E(v_{it}|\boldsymbol{x}_{it}, \boldsymbol{z}_{it}) = 0, \tag{3.6}$$

where

$$d_{it} = \begin{cases} 1 & \text{if} \quad d_{it}^* > 0, \\ 0 & \text{if} \quad d_{it}^* \leq 0. \end{cases} \tag{3.7}$$

如果个体接受处理是随机的，那么就不会存在样本选择效应。此时可以通过回归估计$g_1(x)$和 $g_0(x)$，此时你无需担心“基于可观测变量的选择”问题。假设采用线性回归模型，使用接受处理的个体样本，对自变量 x 回归即可得到 $g_1(x)$；同理，使用未接受处理的个体数据回归，可获得$g_0(x)$。若需控制“基于可观测变量的选择”问题，也可以通过外推法实现。

Then, conditional on $\boldsymbol{x}_{it}$ and d_{it}, the expected value of ε_{it}^j could be either

$$E(\varepsilon_{it}^j|\boldsymbol{x}_{it}, d_{it}) = 0, \quad j = 0, 1. \tag{3.8}$$

or

$$E(\varepsilon_{it}^j|\boldsymbol{x}_{it}, d_{it}) \neq 0, \quad j = 0, 1. \tag{3.9}$$

When (3.8) holds, i.e. $f(\varepsilon^1, \varepsilon^0, v) = f(\varepsilon^1, \varepsilon^0)f(v)$, models (3.1) and (3.2) are typically called the two part model. If the conditional mean function, $g_1(\boldsymbol{x})$ and $g_0(\boldsymbol{x})$, are known, regression methods can be applied to obtain consistent estimator of their parameters. When the conditional mean functions are unknown, nonparametric methods can be applied to identify $g_1(\boldsymbol{x})$ and $g_0(\boldsymbol{x})$ (e.g. Li and Racine (2000)).

如果你已知 $g_1(x)$和 $g_0(x)$的函数形式，你可以直接在相同的x上外推估计 $g_1(x)$和 $g_0(x)$。如果结果方程中的随机误差项ϵ_{it}与参与方程中的随机误差项v_{it}存在相关性，就会产生样本选择效应。在这种情况下，假定实际的 y_{it}^1等于给定条件样本 x_{it}和 $d_{it}=1$的条件期望加上随机误差。若你认为$g_1(x)$的形式为 $x'\beta_1$，而$g_0(x)$为 $x'\beta_0$，则你观察到的y_{it}^1等于$x_{it}'\beta_1+\gamma^1(z_{it})$加上与$x$和 d_{it}不相关的随机误差项 η_{it}^1。其中，$\gamma^1(z_{it})$是在 $v_{it}>-h(z_{it})$时 ϵ_{it}^1的条件期望。类似地，对y_{it}^0，则是在$v_{it}<-h(z_{it})$时 ϵ_{it}^0的条件期望。如果你采用完全参数化的设定，则关于可观测和不可观测变量选择的问题较为简单，计算也很直接，通常采用极大似然估计。

If (3.9) holds, then $f(\varepsilon^1, \varepsilon^0|v) \neq f(\varepsilon^1, \varepsilon^0)$. Model (3.1), (3.2), (3.6), (3.7) together with $f(\varepsilon^1, \varepsilon^0, v)$ is typically referred as sample selection models (e.g. Heckman (1979) and the observed data are subject to *selection on unobservables,* (e.g. Heckman and Vytlacil (2001, 2005)).

If $g_1(\boldsymbol{x}), g_0(\boldsymbol{x})$ and $f(\varepsilon^1, \varepsilon^0, v)$ are known, then observed y_{it}^1 or y_{it}^0 under the assumption that $g_1(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}_1$ and $g_0(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}_0$ takes the form

$$\begin{aligned} y_{it}^1 &= E(y_{it}^1|\boldsymbol{x}_{it}, d_{it} = 1) + \eta_{it}^1 \\ &= \boldsymbol{x}_{it}'\boldsymbol{\beta}_1 + \gamma^1(\boldsymbol{z}_{it}) + \eta_{it}^1, \end{aligned} \tag{3.10}$$

and

$$\begin{aligned} y_{it}^0 &= E(y_{it}^0|\boldsymbol{x}_{it}, d_{it} = 0) + \eta_{it}^0, \\ &= \boldsymbol{x}_{it}'\boldsymbol{\beta}_0 + \gamma^0(\boldsymbol{z}_{it}) + \eta_{it}^0, \end{aligned} \tag{3.11}$$

然而，误差项的概率分布是未知的，你无法知道γ^1 和 γ^0的具体形式。在这种情形下，Robinson（1988）提出了一种方法来处理未知的γ^1 和 γ^0。我们知道$E(y^1|\mathrm{z})=E(x|\mathrm{z})'\beta_1+\gamma^1(z)$，Robinson（1988）建议直接计算$E(y^1|\mathrm{z})$，从而得到$y-E(y^1|\mathrm{z})$等于$(x-E(x|\mathrm{z}))'\beta_1$加上误差项$\eta^1$。根据构造，该误差项与$x-E(x|\mathrm{z})$不相关。因此，可以用参数方法分别估计$E(y^1|\mathrm{z})$和$E(x|\mathrm{z})$。只需将$y-E(y^1|\mathrm{z})$对$x-E(x|\mathrm{z})$进行回归，即可得到$\beta_1$的一致估计，$\beta_0$同理。Powell（2000）进一步将 Robinson（1988）的方法推广，建议采用配对差分估计以消除样本选择偏误。这是横截面数据分析中的基本方法。但这一套方法的前提假设是，所有遗漏因素的随机误差项与协变量x不相关。如果该假设不成立，这些方法将失效。

对于横截面数据，目前尚未见到文献讨论如何在样本选择模型中控制x和ϵ之间的相关性。但在面板数据中可以放宽这一假设。具体来说，可以将ϵ分解为两部分：一部分允许其与x相关，另一部分则假设与x无关。这样，通过面板数据，可以用传统的个体固定效应（个体特有的、不随时间变化的效应）和时间固定效应（同一时间内所有个体都受相同影响，但该影响随时间变化的效应）来表示与x相关的主导因素。Honoré (1992)、Honoré 和 Kyriazidou (2000)、Kyriazidou (1997, 2001)等学者探讨了如何控制遗漏因素之间的相关性，并在没有时点特异性效应假设下，将解释变量纳入模型，这是一种解决方案。

另一种方法考虑如果存在影响结果的其他因素，这些因素不仅在个体之间不同，而且随时间变化。在这种情况下，传统方法无法处理。而因子模型方法可以尝试捕捉这些与解释变量相关且随个体和时间变化的因素的影响，该方法将其分解为在特定时点影响所有个体的共同因子的乘积。尽管这些因子影响所有个体，不同个体所受的影响却不尽相同。这是一种比较流行的方法，用以通过简单模型刻画个体之间及随时间的基本差异。在特定时点存在r个共同因子，它们作用于所有横截面单元，但影响程度不同。Kong 和 Hsiao（2024）研究了当遗漏变量以因子结构表现时的情形。然而，所有这类模型都面临一个问题，即需要进行差分以消除样本选择效应，这意味着例如截距等某些系数无法被估计。我们正在研究该情形，尝试进行修正，目前已取得一些成果。

还有一种方法假设不存在样本选择效应，并且对函数形式g^1和g^0一无所知。这被称为非参数方法。有大量的文献，总结了参数方法、半参数方法以及非参数方法各自的基本优缺点。

Table 1. Advantages and Disadvantages –
Parametric, Semiparametric and Nonparametric Approaches

	Advantage	Disadvantage
Parametric Approach	Simultaneously control the selection of observables and unobservables, issues. Estimate the average treatment effects (ATE) and the impact of each factor. Can obtain efficient estimation of the parameters of the conditional mean function.	Specification of the conditional mean function and the probability distribution of the impact of omitted factors.
Semiparametric Approach	Simultaneously control the selection of observables and unobservables issues. Estimate the impact of most (or some) factors on the outcomes and ATE (in some cases). No need to specify the probability distribution of the impact of omitted factors.	Specification of the conditional mean function. Estimates, although may achieve the same speed of convergence as the parametric approach, it is less efficient.
Nonparametric Approach	No need to specify the conditional mean function and the probability distribution of the impact of omitted factors.	Unconfoundedness is the maintained hypothesis. Curse of dimensionality

我认为参数方法的优点在于，它可以同时控制可观测和不可观测变量的选择，并估计平均处理效应及各因素的影响，同时估计渐近有效。然而，这种方法的缺点在于，必须事先知晓函数形式$g_1(x)$ 和$g_0(x)$，还需了解 ϵ^1、 ϵ^0和v的分布，这对先验知识提出了较高要求。

对于半参数方法，其优势在于能够同时控制可观测因素和不可观测因素的选择带来的影响，而且无需事先知道 ϵ^1和 ϵ^0的联合概率分布，这是半参数方法最大的优点。然而，其劣势在于你仍然需要对条件均值函数g^1和g^0进行设定。此外，这种方法的估计效率较低，且你所关注

的所有系数未必都能被估计出来。在处理效应的测量中，这一点尤为关键，因为此时你确实需要掌握模型中所有参数的信息。

非参数方法的优点在于，无需事先设定条件均值函数或误差项的概率分布函数。然而，其缺点在于无样本选择假设不可检验。纯粹从经济学建模的视角来看，在处理效应的衡量上，各种方法之间并无明显的优劣之分，因为每种方法都具有自身的优势与不足。

非因果方法

正如我在开头所论述的，如果我只关心处理效应的测量，那么我可以很容易地找到一种简单的方法，通过我所称的线性投影（linear projection）方法，无论有怎样的结果变量，都能提供良好的预测。我总可以将 y 关于某些变量的条件均值表示出来，并加上误差项。按照构造，误差项与这些条件变量正交。我要做的只是选择那些不受个体是否接受处理影响的条件变量z_{it}，从而z_{it}的条件密度函数$f(z_{it}|d_{it})$等于其无条件密度函数$f(z_{it})$。

那么，如何预测没有接受处理时的结果呢？我只需给定z_{it}进行预测即可。在这种情况下，面板数据方法（panel data approach）预测反事实的思路，就是找到 z_{it}的最佳函数形式。例如，在 Hsiao 等（2012）的早期工作中，我尝试将y_{it}^1表示为截距加上z_{it}的线性函数。z_{it}的维度可能非常高时，我只需选择z_{it}的一个子集。但无论线性投影关系如何，误差项都是正交的。通过线性回归就可以得到系数，这就是一种方法。

另一种方法是合成控制法（synthetic control method）。该方法试图用其他个体在时点t的结果来预测个体i在同一时点的结果。但合成控制法与面板数据方法存在差异。合成控制法对系数有严格的限制：它要求将其他单位的结果做线性组合以预测目标单位的结果，并且限制系数权重不能为负，同时所有权重之和必须等于 1。然而，如果仅将其视为一种预测，则没有必要引入这些额外假设。已有文献对各种方法的预测准确性进行了比较。然而，一切只能通过样本模拟来实现。如果将实际结果作为论据，那么就难以得出有效结论，因为实际结果尚未发生，亦无法确定。

我想以香港与中国内地签署的《更紧密经贸关系安排》（Closer Economic Partnership Agreement, CEPA）为例。如您所知，香港于 1842 年被割让给英国，并于 1997 年回归中国。主权回归后，香港经历了亚洲金融危机。2003 年，内地与香港签订了 CEPA，该协议有些类似于欧洲的共同市场——香港出口到内地的贸易可以免征关税，专业人士也能在两地之间自由流动。因此，我们希望评估 CEPA 对香港人均 GDP 增长率的影响。在这种情况下，我们掌握了 2004 年以前香港人均 GDP 增长率的数据，但 2004 年以后，我们只能获得香港在 CEPA 框架下的人均 GDP 增长率，而无法知道如果没有签订 CEPA，香港人均 GDP 增长率会如何变化。

在这种情况下，需要构建反事实以评估 CEPA 对香港人均 GDP 增长率的影响。我们使用 2004 年前的季度数据，将香港对奥地利、意大利、韩国、墨西哥、挪威和新加坡进行回归分析，得到了回归系数，标准误差，t统计量，以及R方。接着，我用这个回归系数乘以 2004 年以后奥地利、意大利、韩国、墨西哥、挪威和新加坡的人均 GDP 增长率，构建了一个假设情景，即如果没有 CEPA，香港的人均 GDP 增长率会如何发展。下图左侧显示的是实施 CEPA 后香港实际的人均 GDP 增长率。实际情况与该假设之间的差距在第一个季度大约为 2.7%。

Table 9.1. AICC Selected Model Using Data for the Period 1993Q1 – 2003Q4

	Beta	Std	T
Constant	-0.0019	0.0037	-0.524
Austria	-1.0116	0.1682	-6.0128
Italy	-0.3177	0.1591	-1.9971
Korea	0.3447	0.0469	7.3506
Mexico	0.3129	0.051	6.1335
Norway	0.3222	0.0538	5.9912
Singapore	0.1845	0.0546	3.3812
$R^2 = 0.931$			
AICC = -378.9427			

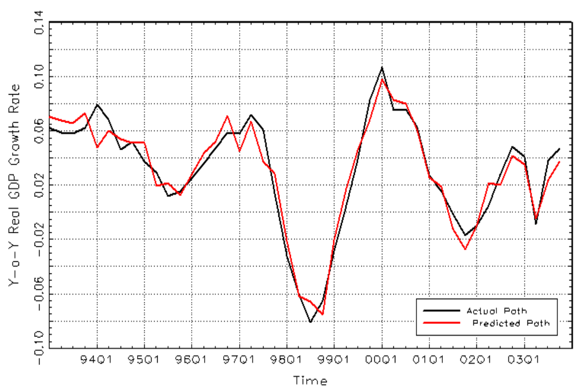
Source: Hsiao et. al. (2012, Table 20)

Table 9.2. Treatment Effect for Economic Integration 2004Q1 – 2008Q1 Based on AICC Selected Model

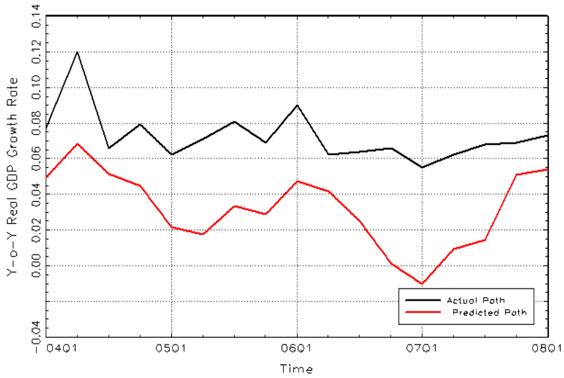
	Actual	Control	Treatment
Q1-2004	0.077	0.0493	0.0277
Q2-2004	0.12	0.0686	0.0514
Q3-2004	0.066	0.0515	0.0145
Q4-2004	0.079	0.0446	0.0344
Q1-2005	0.062	0.0217	0.0403
Q2-2005	0.071	0.0177	0.0533
Q3-2005	0.081	0.0333	0.0477
Q4-2005	0.069	0.029	0.04
Q1-2006	0.09	0.0471	0.0429
Q2-2006	0.062	0.0417	0.0203
Q3-2006	0.064	0.025	0.039
Q4-2006	0.066	0.0009	0.0651
Q1-2007	0.055	-0.0101	0.0651
Q2-2007	0.062	0.0092	0.0528
Q3-2007	0.068	0.0143	0.0537
Q4-2007	0.069	0.0508	0.0182
Q1-2008	0.073	0.0538	0.0192
MEAN	0.0726	0.0323	0.0403
STD	0.0149	0.0213	0.016
T	4.8814	1.5132	2.5134

Source: Hsiao et. al. (2012, Table 21)

从图中可以直观地看出这一点。黑色曲线表示香港 2004 年之前的人均 GDP 增长率的实际值，红色曲线则为拟合值。这属于样本内拟合。



这是 2004 年后的预测结果。黑色曲线表示香港实际人均 GDP 增长率。红色曲线表示在没有 CEPA 情况下预测的香港人均 GDP 增长率。可以看出，CEPA 对香港人均 GDP 增长率存在正向影响。



另一种方法是因子方法（factor approach），试图用一个简单的模型来捕捉个体之间和随时间变化的差异，该方法将观测结果分解为r个共同因子f_t及其对每个个体i单元的影响λ_i的乘积，再加上误差项。这是一种常见的用于总结个体之间以及时间上的变异性的方法。你可以观测到y_{it}，但无法直接观测到λ_i和f_t。要识别λ_i，我需要时间序列维度；要识别f_t，则需要横截面维度。因此，作为估计方法，我个人认为因子方法非常好，它能够总结个体在不同时期内的变化信息，但作为预测的方法并不是很好。

周前坤和我最近发表了一项研究，比较了因子方法与线性投影方法在预测性能上的表现。简要总结我们的结果如下：当N和T的维度固定时，无法得出明确结论。但如果截面维度较大、时间序列维度固定，那么从均方预测误差（MSPE）的角度来看，线性投影方法优于因子方法。如果时间序列维度较大、N固定，同样从预测精度来看，线性投影方法更为准确。类似地，当截面维度和时间序列维度都较大时，线性投影方法的 MSPE 表现也占优势。

Table 2. Mean Square Prediction Error (MSPE) -Linear Projection (LP) vs. Factor Models (FB)

	N Fixed	$N \to \infty$
T fixed	No definite conclusion. The MSPE depends on realized $\boldsymbol{y}_{it}$ and $\boldsymbol{x}_{it}$.	MSPE (modified LP) ≤ MSPE (FB)
$T \to \infty$	MSPE (LP) ≤ MSPE (FB)	MSPE (LP) ≤ MSPE (FB)

多个处理个体

另一个问题是多个个体接受处理。在这种情况下，我可以为每个个体生成预测，最后取平均值。如果接受处理的个体数量较多，另一种方法是将所有这些个体加总为一个单位，并采用我之前讨论的方法。然而，具体的应用过程及该问题的条件尚未被深入探讨。例如如果所有个体可以被视为同质接受处理，简单地取平均值就是一个较好的估计方法。但当个体存在差异时，如何将个体进行加总就成为一个问题，在这个问题上，现有文献并未给予充分关注。加总时应

当侧重个体间的异质性，还是应强调个体之间的共同行为趋势？不同的侧重点将导致不同的加总方法。我与沈艳、周前坤曾简要涉及过这一问题，但该问题尚未被深入讨论。

另一个方面是关于处理效应的度量。如果不希望使用平均处理效应作为度量标准，如何评估特定政策的有效性？对此，Maasoumi 和 Wang（2019, 2020）提出了一种方法。假设有一个关于处理效应在不同结果变量y取值处的效用函数，那么可以根据不同政策下结果的分布进行积分，从而判断政策是否有效。

我还想补充一点，由于面板数据可以展示个体间影响的异质性，因此可以分析不同结果之间的差异，找出导致个体间差异的关键因素。例如，有位厦门大学的博士研究了高速铁路对不同地区的影响。在下图中，横轴表示年份，纵轴表示影响值。在北京至南京段，廊坊的影响一开始为零，但随着时间推移，影响逐渐减弱并变为负值。而徐州的影响则始终为正，并且逐渐上升。那么，为何廊坊与徐州的影响差异如此明显？你可以通过回归分析的方法来探究导致这些差异的原因。

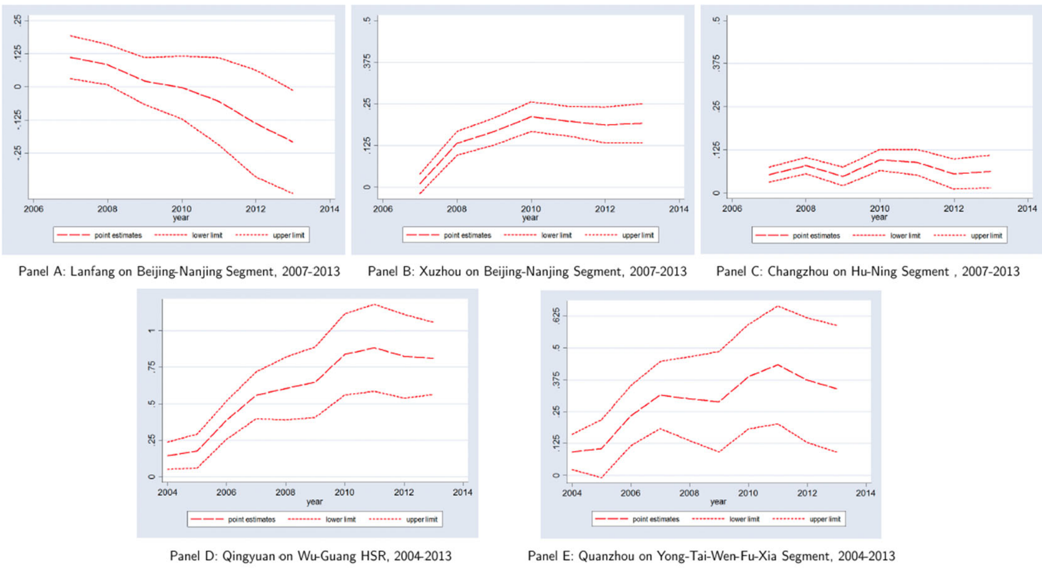


Figure 2: Treatment effects in per capita real GDP for the five HSR cities

结语

我在此总结处理效应的测量问题。如果采用传统经济学模型的方法，则必须能够识别模型中的所有参数，但遗憾的是，除参数化方法外，其他方法通常无法做到这一点。如果只是为了预测，那么线性预测方法相对容易实施，并且能够带来良好的预测效果。

另外我想讨论的一点是，所有这些测量方法都基于一个假设：即只有一个政策。你所观察到的数据，只是在这个政策有或无的情况下产生的。然而，在这个时间段内实际上可能存在两种不同的政策。那么，如何区分这两个政策的影响呢？在这种情况下，可能需要结合多种方法，分别对每种政策的处理效应进行独立估计。

最后，我想说，如果只是进行预测，那么就无法进行政策模拟。评估和解释“如果”情形，需要有经济模型。比如，我的同事 Pesaran 与他的学生合作评估了新冠疫情暴发期间，各国实

施不同控制政策的影响。他们基于网络 SIR 模型，构建了计量经济学模型，分析了疫情通过经济网络在一个月内的影响。

他们想要观察在实施社交距离政策后，会出现什么样的结果。下图中左上子图蓝色区间展示了德国实际实施社交距离政策后的影响，以及在政策生效期间仍存在的潜在感染者。如果德国将社交距离政策延迟一周实施，影响则如左上子图红色区间所示。对于阳性病例数也是同样的道理，中上子图展示了在社交距离限制下的实际结果，以及如果政策延迟实施所导致的影响。同样地，他们还基于模型估计的参数，分析了英国的情况，例如如果英国能提前一周实施控制措施，疫情的发展是否会有不同的结果。正因为有经济模型作为工具，才能进行这样详细的分析，否则无法得出这些结论，这也是模型的另一大价值。

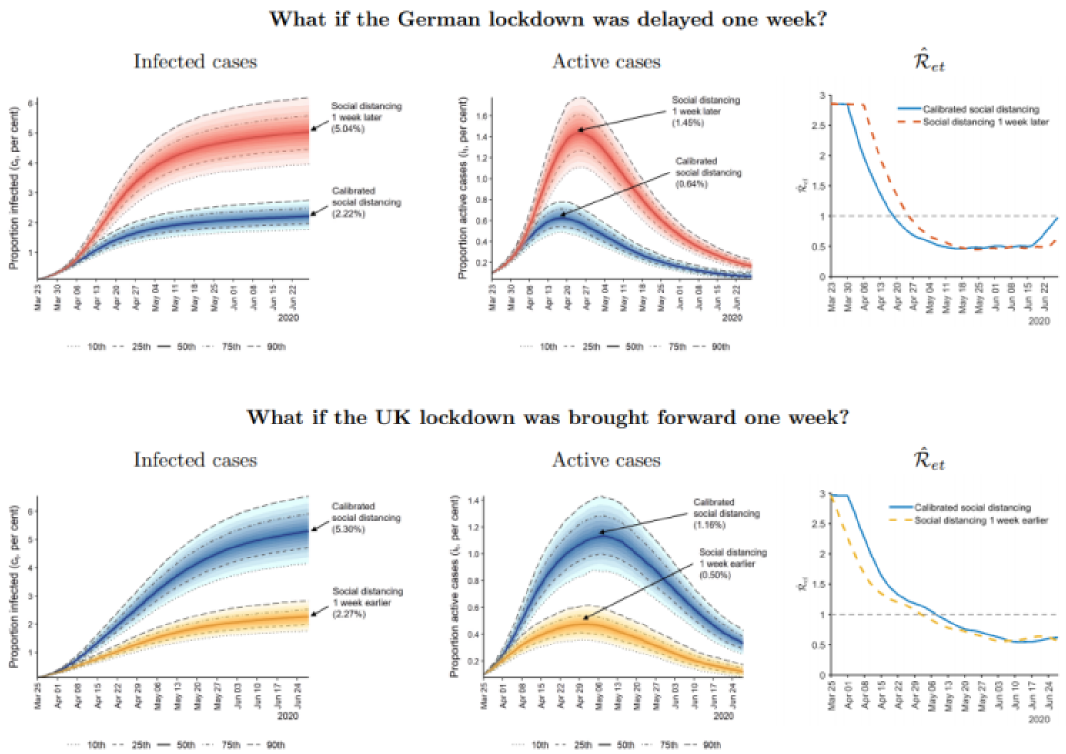


Figure 12.7 (a) Counterfactual number of infected and active cases for Germany and UK

但需要注意，在使用经济模型时我之前讨论的方法都是单方程方法，即使经济模型本身是一个局部均衡模型，不同因素的局部变化也可能产生不同影响，这便涉及联立方程系统。据我所知，目前尚无文献在逻辑上基于联立方程来衡量潜在的处理效应。例如，中国宣布利率降低，利率的变化既可能影响消费，也可能影响投资，进而影响中国经济的总体增速。而单方程或 AI 方法通常只考虑利率下降对某一结果变量的影响，而未同时衡量其对投资、消费以及最终经济表现的综合作用，这正是现有方法中的一个不足之处。

我就讲到这里，感谢各位的关注！