2014 年 9 月
第 33 卷 第 5 期
数理统计与管理
Journal of Applied Statistics and Management
Sep. 2014
Vol. 33 No. 5

文章编号：1002–1566 (2014) 05–0830–12
DOI: 10.13860/j.cnki.sltj–20140922–024

Hausman 检验统计量有效性的 Monte Carlo 模拟分析

连玉君[1]　王闻达[2]　叶汝财[1]

(1. 中山大学岭南学院，广东 广州 510275; 2. 中国人民大学汉青经济与金融高级研究院，北京 100872)

摘要：在大样本下，Hausman 检验统计量渐进地服从卡方分布，但在有限样本分析中却经常出现负值。对此，文献中并未形成一致看法。本文采用 Monte Carlo 模拟，从小样本偏误、内生性偏误、序列相关、非平行面板等角度，研究了 Hausman 检验统计量的小样本性质。结果表明，内生性问题 (解释变量与个体效应相关) 是导致 Hausman 统计量出现负值的主要原因。进一步分析表明，修正后的 Hausman 统计量，以及过度识别检验方法能够很好地克服上述缺陷，且具有很好的有限样本性质。

关键词：固定效应模型；随机效应模型；面板数据；Hausman 检验；Monte Carlo 模拟分析
中图分类号：F064.1, O212　　**文献标识码**：A

The Efficiency of Hausman Test Statistics: A Monte-Carlo Investigation

LIAN Yu-jun[1]　WANG Wen-da[2]　YE Ru-cai[1]

(1. Sun-Yat-Sen University, Guangdong Guangzhou 510275, China, 2. Renmin University of China, Beijing, 100872)

Abstract: In large sample, the Hausman test statistic is distributed as χ^2 asymptotically. However, in small sample, its application quite often leads to negative values. Previous literature is substantially inconsistent with this issue. This paper investigates the small sample properties of the Hausman test statistics using Monte-carlo simulation. The possible biases considered including small sample bias, endogeneity, serial correlation, unbalanced panel data. The results show that the main source of a negative Hausman statistic is the endogeneity problem, i.e., the orthogonality condition between the regressors and the individual effects fails. Further results show that, the modified Hausman statistics and the overidentify restrictions test can be used to avoid the negative values problem. Moreover, these additional test statistics perform well in finite sample.

Key words: fixed effects model, random effects model, panel data, hausman test, monte-carlo simulation

0 引言

对于面板数据而言，最为常用的模型是固定效应模型 (简称 FE) 和随机效应模型 (简称 RE)。二者的区别在于对个体效应的处理方式不同。FE 模型假设个体效应是非随机变量，因

收稿日期: 2011 年 7 月 16 日　　**收到修改稿日期**: 2012 年 12 月 10 日
基金项目: 国家自然科学基金 (71002056)、教育部人文社会科学研究基金项目 (09YJC790269) 资助。

此可以视为解释变量的一部分；RE 模型则假设个体效应是随机的，并将其视为干扰项的一部分。相对而言，FE 的假设条件比较宽松，而 RE 的假设条件则比较严格 (假设个体效应与所有的解释变量都不相关，即模型设定不存在内生性问题)。因此，在实证分析过程中，多数学者都倾向于使用 FE 模型。然而，在个体效应与解释变量不相关的情况下，RE 模型具有自身的独特优势。一方面，相对于传统的 Pooled OLS 模型，RE 模型的设定只增加了一个参数，即个体效应的标准差，而 FE 模型则需要额外估计 $N-1$ 个参数，这使得 RE 估计量更为有效。另一方面，在有些分析中，我们会非常关注那些不随时间变化的变量 (如性别、种族、出生地等) 对某个变量的影响。在 FE 模型中，由于这些变量都与反映个体效应的虚拟变量完全共线性，会存在无法识别的问题，而在 RE 模型中，由于个体效应是随机干扰项的一部分，这些变量是可以识别和估计的。

因此，在实证分析过程中，如何选择 FE 和 RE 模型便成为一个非常重要的问题。虽然计量经济学家提出了多种方法来检验二者的适用性，但并未得到一致且足以令人信服的结论。相对而言，Hausman 检验应用最为广泛 [1]，许多实证研究者都采用这种方法选择相应的面板模型，如王敏等 [2] 和柯孔林 [3]。然而，该统计量的一个主要缺陷在于，虽然从理论上讲，检验统计量渐进服从于卡方分布，但实际计算出来的统计值 (以下简称 “h 统计量”) 经常出现小于零的现象。

在这种情况下，探求 “h 统计量” 为负的根源就显得尤为必要，否则将无法在 FE 和 RE 估计量之间做出取舍。遗憾的是，文献中并未对此问题形成一致观点。一些主流的计量经济学教科书都倾向于将这种异常的情况归因于小样本偏误 [4–6]，学界广泛使用的计量软件 Stata 便采用了这种观点，其使用手册认为 “h 统计量” 为负的情况时有发生，并建议此时应该接受原假设 [7:642]。并认为可以将负的 h 统计值视为零，进而进行后续统计推断。这意味着，当出现 h 统计量为负的情形时，应该接受原假设而选用 RE 估计量。然而，在近期的研究中，Schreiber[8]、Magazzini 和 Calzolari[9] 等学者却提出的截然相反的观点，他们的理论分析表明，h 统计量为负事实上恰恰表明原假设不合理，此时应该选用 FE 模型。更为重要的是，这种状况不仅会发生在小样本中，对于大样本而言，这种情况在特定的条件下也会发生。上述研究多集中于大样本，试图从理论上探求 h 统计量出现负值的原因，但对于小样本情形下 h 统计量的表现却鲜有文献提及，而后者恰恰是实证分析中真正需要处理的问题。

国内学者尚未从计量理论角度研究 h 统计量的性质，但在各个领域的实证研究中，当面对 h 统计量为负这一问题时，学者们普遍存在困惑，以至于采取了方式各异甚至截然相反的处理措施。部分学者遵循了传统教科书及 Stata 软件手册的建议，将 h 统计量出现负值视为不能拒绝原假设的信号，并据此选择随机效应模型，比如金煜等 [10:87]，以及薄文广 [11:7]；有些学者则认为这是一种异常情况，将其归因为模型设定偏误问题 [12:110]，进而从模型设定角度进行更为深入的分析；有的学者在处理高频数据的时候也出现了 Hausman 检验出现负值的情况，如沈根祥 [13] 认为在抽样频率很高的时候，微观结构的噪音是影响 Hausman 检验出现负值的一个主要原因；另有部分学者则出于稳健性考虑直接选择了固定效应模型，如张卉等 [14:23]、李婉丽和吕怀立 [15:87]；还有学者选择放弃 Hausman 检验另寻办法 [16:38]，或者直接从模型的经济意义上来进行判断 [17:64]。这意味着，对 h 统计量的小样本性质，尤其是其出现负值的根源进行研究显得尤为必要。

显然，从理论上研究 h 统计量的小样本性质并非易事。为此，本文尝试采用 Monte Carlo 模拟方法进行研究，主要解决如下两个问题：其一，导致 h 统计量出现负值的根源是什么？其二，当 h 统计量出现负值时该如何解决？

我们考察了内生性偏误、序列相关偏误、非平行面板等对 h 统计量小样本性质的影响。结果表明，内生性问题是导致 h 统计量出现负值的主要原因。换言之，实证分析中出现 h 统计量为负，主要是因为 RE 模型的原假设无法满足所致。因此，在遇到 h 统计量小于 0 的情况时，应当重新考虑模型的设定，尤其是注意个体效应和解释变量是否相关。此时，作为替代方法，可以采用 Hayashi[18] 建议的修正 h 统计量，或 Arellano[19] 提出的过度识别检验，二者都能有效避免统计量为负值的情形，且在小样本下有很好的统计性质。

后文结构安排如下：第 1 节为研究设计，介绍模拟数据的生成过程；第 2 节呈现 Monte Carlo 模拟结果及分析；第 3 节是扩展与建议，分析了 Hausman 检验替代方法的小样本性质；最后是结论。

1 研究设计

1.1 传统的 Hausman 检验方法

给定如下面板数据模型：

$$y_{it} = \mathbf{x}_{it}\beta + \mu_i + \varepsilon_{it}, \qquad i = 1, 2, \cdots, N, \quad t = 1, 2, \cdots, T, \tag{1}$$

其中，y_{it} 是被解释变量；$\mathbf{x}_{it}$ 是 K 个解释变量构成的向量，i 和 t 分别表示截面和时间标识。μ_i 为不可观测的个体效应，ε_{it} 为随机干扰项。对此模型有两种估计方法：FE 和 RE。二者所需满足的假设条件如下 [20]：

$$\begin{aligned}
&H_{FE}^1:\ E(\varepsilon_{it}|\mathbf{x}_i, \mu_i) = 0, \qquad \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2),\\
&H_{RE}^1:\ E(\varepsilon_{it}|\mathbf{x}_i, \mu_i) = 0, \qquad \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2), \quad \mu_i N(0, \sigma_\mu^2),\\
&H_{RE}^2:\ E(\mu_i|\mathbf{x}_i) = E(\mu_i) = 0.
\end{aligned}$$

对比 H_{FE}^2 和 H_{RE}^1 可知，FE 和 RE 都要求干扰项 ε_{it} 与解释变量 $\mathbf{x}_{it}$ 不相关，即二者的模型设定都不存在一般意义上的内生性问题。二者的区别主要表现为 H_{RE}^2 的差异上。由于 FE 假设个体效应 μ_i 是固定的，可以视为解释变量的一部分，而 RE 则假设 μ_i 是干扰项的一部分，因此，在 RE 模型的设定中，必须进一步要求 μ_i 与 $\mathbf{x}_{it}$ 不相关，即模型设定不存在面板特定的内生性问题。

在满足上述假设的前提下，要得到 FE 估计量 $\hat\beta_{FE}$，可先通过对模型 (1) 进行“组内差分变换”(within group difference) 以去除个体效应 μ_i，进而对如下模型执行 OLS 估计：

$$(y_{it} - \overline{y}_i) = (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\beta + (\varepsilon_{it} - \overline{\varepsilon}), \tag{2}$$

其中，$\overline{y}_i = \sum_{t=1}^{T_i} y_{it}/T$，$T_i$ 表示第 i 个截面包含的时间跨度，$\overline{\mathbf{x}}_i$ 和 $\overline{\varepsilon}_i$ 的定义与此相似。对于 RE 估计量 $\hat\beta_{RE}$，给定 $\hat\sigma_\varepsilon^2$ 和 σ_μ^2 的一致估计量 σ_ε^2 和 σ_μ^2(对于 $\hat\sigma_\varepsilon^2$ 和 $\hat\sigma_\mu^2$ 的估计方法，参见 Baltagi[5:19] 或 StataCorp[21:464])，需要先对模型进行“GLS 变换”：

$$(y_{it} - \hat\theta_i\overline{y}_i) = (\mathbf{x}_{it} - \hat\theta_i\overline{\mathbf{x}}_i)\beta + \big((1-\hat\theta_i)\mu_i + (\varepsilon_{it} - \hat\theta_i\overline{\varepsilon}_i)\big), \tag{3}$$

其中，$\hat\theta_i = 1 - \sqrt{\hat\sigma_\varepsilon^2/(T_i\hat\sigma_\mu^2 + \hat\sigma_\varepsilon^2)}$。对模型 (3) 执行 OLS，即可得到 $\hat\beta_{RE}$。我们可采用矩阵形式将 $\hat\beta_{FE}$ 和 $\hat\beta_{RE}$ 以及二者的方差 - 协方差矩阵表示如下：

$$\begin{aligned}
\hat\beta_{FE} &= (\ddot{\mathbf{X}}'\ddot{\mathbf{X}})^{-1}\ddot{\mathbf{X}}'\ddot{y}, \qquad \hat V_{FE} = \hat\sigma_\varepsilon^2(\ddot{\mathbf{X}}'\ddot{\mathbf{X}})^{-1},\\
\hat\beta_{RE} &= (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{y}, \qquad \hat V_{RE} = \hat\sigma_\varepsilon^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1},
\end{aligned} \tag{4}$$

其中，对于 $\hat{\beta}_{FE}$，$\check{\mathbf{X}}=(\check{\mathbf{X}}_1',\check{\mathbf{X}}_2',\cdots,\check{\mathbf{X}}_N')'$，$\check{y}=(\check{y}_1',\check{y}_2',\cdots,\check{y}_N')'$，$\check{z}_{it}=(z_{it}-\bar{z}_i)$ (z_{it} 表示 x 或 y_{it})。对于 $\hat{\beta}_{RE}$，只需定义 $\tilde{z}_{it}=(z_{it}-\hat{\theta}_i\bar{z}_i)$，即可按照上述表述方式定义 $\tilde{\mathbf{X}}$ 和 $\tilde{\mathbf{y}}$。

Hausman[1] 检验的基本思想在于：若原假设 H_{RE}^2 成立，则 $\hat{\beta}_{FE}$ 和 $\hat{\beta}_{RE}$ 都是 β 的一致估计量，虽然 $\hat{\beta}_{RE}$ 更为有效，但二者不会表现出系统性差异；然而，若原假设 H_{RE}^2 不成立，则 $\hat{\beta}_{FE}$ 仍然是 β 的一致估计量，而 $\hat{\beta}_{RE}$ 则是有偏的，此时二者将表现出系统性差异。基于此，Hausman[1] 构造了如下 h 统计量：

$$h=(\hat{\beta}_{RE}-\hat{\beta}_{FE})'(\hat{V}_{FE}-\hat{V}_{RE})^{-1}(\hat{\beta}_{RE}-\hat{\beta}_{FE}). \tag{5}$$

在原假设 H_{RE}^2 下，h 统计量渐进地服从卡方分布，自由度为 K，即 $h\sim\chi^2(K)$。

1.2 修正后的 Hausman 统计量

鉴于式 (4) 中 h 统计量在小样本中经常出现负值的局限，Cameron 和 Trivedi[22:260] 建议采用修正后的 h 统计量。他认为，修正后 h 统计量出现负值的概率会明显降低。修正方法是在计算 h 统计量时对 $\hat{V}_{RE}$ 和 $\hat{V}_{FE}$ 进行调整。基于 $\hat{V}_{FE}$ 调整得到的统计量为 [20:290]：

$$h_{FE}=(\hat{\beta}_{RE}-\hat{\beta}_{FE})'\left(\hat{V}_{FE}-\frac{RMSE_{FE}^2}{RMSE_{RE}^2}\hat{V}_{RE}\right)^{-1}(\hat{\beta}_{RE}-\hat{\beta}_{FE}), \tag{6}$$

即，保持 $\hat{V}_{FE}$ 不变而调整 $\hat{V}_{RE}$，其中 $RMSE_{FE}$ 和 $RMSE_{RE}$ 分别是 FE 模型和 RE 模型的均方根误差。基于 $\hat{V}_{RE}$ 的调整方法类似，相应的统计量为：

$$h_{RE}=(\hat{\beta}_{RE}-\hat{\beta}_{FE})'\left(\frac{RMSE_{RE}^2}{RMSE_{FE}^2}\hat{V}_{FE}-\hat{V}_{RE}\right)^{-1}(\hat{\beta}_{RE}-\hat{\beta}_{FE}). \tag{7}$$

1.3 模拟数据的生成过程 (DGP)

通过上述分析可以推测，Hausman 检验的原假设 H_{RE}^2 无法得到满足 (内生性问题) 或许是导致 h 统计量出现负值的主要原因。Schreiber[8]、Magazzini 和 Calzolari[9] 等学者基于大样本的理论分析也支持这种观点。为此，在 Monte Carlo 模拟分析过程中，本文重点探讨内生性问题对 h 统计量小样本性质的影响，同时也可考虑了其他几种潜在的偏误。具体而言，本文拟分析的偏误包括：(B1) 内生性问题：解释变量与个体效应相关，即 $\operatorname{corr}(x_{it},\mu_i)\neq 0$；(B2) 一般意义下的内生性问题：解释变量与扰动项相关，即 $\operatorname{corr}(x_{it},\varepsilon_{it})\neq 0$；(B3) 序列相关；(B4) 非平行面板。

1) 基准模型的数据生成过程。首先，我们设定了一个标准的静态面板模型：

$$y_{it}=\beta x_{it}+\mu_i+\varepsilon_{it}, \qquad x_{it}=\varphi x_{it-1}+\xi_{it}, \tag{8}$$

其中，y_{it} 是被解释变量；x_{it} 是外生的解释变量，$i=1,\cdots,N$，$t=1,\cdots,T$。考虑到在实证分析中解释变量通常存在自相关关系，借鉴 Baltagi et al.[23]，本文将 x 设定为一个 AR(1) 过程；μ_i 为不可观测的个体效应，$\mu_i\sim N(0,1)$；ε_{it} 和 ξ_{it} 是不可观测的随机干扰项，$\varepsilon_{it}\sim N(0,1)$，$\xi_{it}\sim N(0,1)$。在后续模拟分析中，我们都设定 $\beta=0.8$，$\varphi=0.2$。

2) 内生性偏误。在偏误 (B1) 中，为了反映 x_{it} 与 μ_i 之间的相关性，我们借鉴 Magazzini 和 Calzolari[9]，将 x_{it} 的数据生成过程定义如下：

$$x_{it}=0.2x_{it-1}+\theta\mu_i+\sqrt{1-\theta^2}\,v_{it}, \qquad v_{it}\sim N(0,5^2), \tag{9}$$

其中，θ 值反映了内生性偏误的严重程度。在模拟过程中，通过设定不同的 θ $(0<\theta<1)$ 来分析 h 统计量的表现。这里，考虑到当 $\theta \to 1$ 时，x_{it} 主要受个体效应 μ_i 的影响，使其几乎不随时间改变。因此，在采用 FE 模型估计时，式 (2) 的组内差分变换会损失 x_{it} 中包含的多数信息，致使 FE 估计量趋近于 RE 估计量 (Schreiber[8] 对此进行了较为详细的论述，并认为在数据生成过程中，需要适当增大 x_{it} 的时序波动)。同时，这也与实证分析中常见的数据情形不符。因此，本文将干扰项 v_{it} 的标准差设为 5(而非 1) 以增大其波动性，使得即使当 θ 很大时 x_{it} 仍会随时间改变。

3) 一般意义下的内生性偏误。在偏误 (B2) 中，采用如下方式反映 x_{it} 与 ε_{it} 的相关性：

$$\varepsilon_{it}=\gamma x_{it}+w_{it}, \qquad w_{it}\sim N(0,1). \tag{10}$$

这里，参数 γ $(0<\gamma<1)$ 反映了一般意义下内生性偏误的严重程度。需要说明的是，除了 Hausman 检验中明确设定的原假设 H_{RE}^2 外，其实该检验还隐含着 H_{FE}^1 和 H_{RE}^1 两个辅助的假设条件。因此，式 (9) 中设定的一般意义下的内生性偏误，目的在于考察 H_{RE}^2 为真，但 H_{FE}^1 和 H_{RE}^1 不成立时 h 统计量的表现。显然，这种偏误在实证分析中非常普遍，此时无论是 FE 还是 RE 都不再是一致估计量，在这种情况下，FE 和 RE 都是有偏的，而且都倾向于下偏，Magazzini 和 Calzolari[9] 基于大样本的理论分析表明，此时 h 统计量出现负值的概率反而较小。

4) 序列相关。在偏误 (B3) 中，我们将干扰项设定为如下 AR(1) 过程：

$$\varepsilon_{it}=\rho\varepsilon_{it-1}+v_{it}, \qquad v_{it}\sim N(0,1). \tag{11}$$

在模拟分析中，通过选取不同的 ρ 值 $(-1<\rho<1)$ 来考察 h 统计量的表现。之所以考虑序列相关偏误，主要出于两个目的：其一，干扰项的序列相关往往是导致内生性问题的一个潜在原因；其二，虽然这种偏误不会影响 β 的无偏性和一致性，但去会影响其方差协方差矩阵，由式 (4) 可知，这会对 h 统计量的估计值产生影响。

5) 非平行面板。对于偏误 (B4)，在多数实证分析中，由于样本中不断有个体 (个人、企业或地区) 加入或退出，使得最终使用的样本都是非平行面板数据 (unbalanced panel data)。个体的加入或退出可能导致样本选择偏误 [24]，并进而产生内生性偏误 [25]。相对于上文提到的偏误 B1 和 B2，非平行面板导致的偏误更为隐蔽。借鉴 Baltagi 和 Chang[26] 以及 Bruno[27] 的处理方法，本文同样采用 Ahrens 和 Pincus[28] 指数来控制面板的非平行程度，定义如下：

$$\omega=N\Big(\overline{T}\sum_{i=1}^{N}\frac{1}{T_i}\Big)^{-1}, \tag{12}$$

其中，T_i 表示第 i 个截面包含的时间跨度。样本中的总观察值个数为 $n=\sum_{i=1}^{N}T_i$，$\overline{T}=n/N$ 为平均时间跨度。显然，$0<\omega\le 1$。当 $\omega=1$ 时，面板数据是平行的，ω 越接近 0，面板的非平行程度将越严重。

2 Monte Carlo 模拟结果及分析

2.1 传统 Hausman 统计量的统计性质

给定上述模拟数据的生成过程，在 Monte Carlo 模拟分析过程中，本文将模拟次数设定为 1000 次，置信水平为 95%(即，显著性水平为 5%)。

1) 基准模型的模拟结果。在检验上述提及的各种偏误对 h 统计量的影响之前，我们首先针对基准模型 (7)，即不存在任何偏误的模型，模拟分析了传统 "h 统计量" 和修正后的 "h_{RE} 和 h_{FE} 统计量" 的小样本性质，结果呈现于表 1。

表 1　基准模型 Hausman 检验统计量的表现

检验统计量	$N=200, T=10$		$N=1000, T=10$		$N=30, T=30$	
	A	B	A	B	A	B
h	0.00	5.00	0.00	4.70	10.60	6.70
h_{RE}	0.00	4.40	0.00	4.40	0.00	4.90
h_{FE}	0.00	4.40	0.00	4.40	0.00	5.00

注：A 代表检验统计量为负的比例，B 代表在 5% 显著水平下拒绝原假设的百分比。

我们重点考察了三种典型的面板数据结构：$(N,T)=(200,10)$ 和 $(N,T)=(1000,10)$ 是研究企业行为时的典型数据形态，而 $(N,T)=(30,30)$ 则是区域经济增长文献中经常出现的数据形态。从表中 A 栏所示的模拟结果来看，在 $(N,T)=(200,10)$ 和 $(N,T)=(1000,10)$ 这两种面板结构下，h 统计量以及 h_{RE}、h_{FE} 统计量均未出现负值。同时，由 B 栏的结果可知，在 1000 次的模拟中，显著拒绝原假设的比例均不超过事先设定的显著水平 5%。这表明，在模型设定不存在任何偏误的情况下，无论是传统的 h 统计量还是修正后的 h 统计量都具有非常好的小样本性质。

然而，当面板结构为 $(N,T)=(30,30)$ 时，虽然 h_{RE} 和 h_{FE} 统计量的性质并未发生明显变化，但传统的 h 统计量却出现了小于 0 的情形，比例达到了 10.6%。由于基准模型不包含任何设定偏误，那么唯一可能导致 h 统计量为负的原因就是样本太小，这从一定程度上验证了主流教科书中的观点，如 Cameron 和 Trivedi[4]、Baltagi[5]、Hsiao[6]。

表 2　样本数较小时 Hausman 检验统计量的表现

检验统计量	$N=10, T=3$		$N=10, T=30$		$N=10, T=60$	
	A	B	A	B	A	B
h	4.20	8.40	23.50	2.30	26.80	1.70
h_{RE}	0.00	6.70	0.00	7.40	0.00	7.70
h_{FE}	0.00	7.80	0.00	7.70	0.00	7.70

注：A 代表检验统计量为负的比例，B 代表在 5% 显著水平下拒绝原假设的百分比。

为此，本文又进一步设定了 $N=10$，$T\in\{3,30,60\}$ 三种小样本数据，模拟结果呈现于表 2。可以看出，在三种情形中，h 统计量出现负值的概率分别为 4.2%、23.5% 和 26.8%，呈现逐渐增加的趋势。这进一步证实了上述观点，即，h 统计量在样本较小时表现欠佳，且出现负值的概率很高。同时，我们也注意到，在小样本情形下，虽然 h_{RE} 和 h_{FE} 统计量始终未出现负值，但却会在一定程度上过度拒绝原假设。例如，在 $(N,T)=(10,30)$ 面板结构中，h_{RE} 和 h_{FE} 拒绝原假设的比例分别为 7.4% 和 7.7%，均高于我们设定的显著水平 5%。

2) 内生性偏误下的模拟结果。对于内生性偏误 (B1)，类似于表 1，我们同样设定了 $(N,T)=(200,10)$、(1000,10)，(30,30) 三种典型的面板结构，结果呈现于表 3。对于 $(N,T)=(200,10)$，由 A 栏结果可知，当内生性偏误较轻时 $(\theta\le 0.2)$，h 统计量并未出现小于零的情形。然而，随着 θ 值的增大，即 x_{it} 与 μ_i 相关程度的提高，h 统计量出现负值的概率也明显增加，尤其是当 $\theta=0.9$ 时，1000 次 Monte Carlo 模拟中有 942 次 h 统计量都为负值，比例高达 94.2%。从 B 栏的结果来看，虽然在 θ 较小时 $(\theta\le 0.3)$，h 统计量几乎不会出现负值，但此时其检定力

(Power) 却远低于我们设定的置信水平 (95%)。随着 θ 的进一步增大，只有在 $\theta=0.4$ 时，h 统计量的检定力比较接近 95%，而当 $\theta\geq 0.6$ 时，h 统计量的检定力都远低于 95%。

综合 A 栏和 B 栏的结果来看，内生性偏误 (B1) 是导致 h 统计量为负的一个主要原因，在这种情况下，若当 h 统计量出现负值时接受原假设 H_{RE}^2，则其检定力非常低。然而，如果将 h 统计量出现负值归因为内生性偏误，并进而拒绝原假设的话，那么其检定力就等于 A 栏和 B 栏中数值的加总。按此定义，当 $\theta\geq 0.4$ 时，h 统计量的检定力都超过了 95%，尤其是当 $\theta\geq 0.6$ 后，其检定力接近 100%。这为 Schreiber[8]、Magazzini 和 Calzolari[9] 的理论预期提供了直接的经验证据，即，当 h 统计量出现负值时应该拒绝原假设。

表 3　内生性偏误下 Hausman 检验统计量的表现

θ 值	$N=200,T=10$		$N=1000,T=10$		$N=30,T=30$	
	A	B	A	B	A	B
0.1	0.00	15.60	0.00	49.60	16.30	8.40
0.2	0.00	41.40	0.00	97.30	32.20	8.60
0.3	0.30	78.00	0.00	100.00	56.60	11.80
0.4	1.70	94.40	0.00	100.00	76.30	8.50
0.5	8.40	91.30	0.20	99.80	88.80	5.80
0.6	29.80	70.20	16.30	83.70	96.90	2.60
0.7	69.10	30.90	89.20	10.80	99.20	0.80
0.8	95.40	4.60	100.00	0.00	100.00	0.00
0.9	94.20	5.80	100.00	0.00	99.90	0.10

注：A 代表检验统计量为负的比例，B 代表在 5% 显著水平下拒绝原假设的百分比。

进一步分析 $(N,T)=(1000,10)$ 和 $(30,30)$ 两种面板结构，可以发现一些有趣的结论。对于前者而言，只有当 $\theta\geq 0.5$ 时，h 统计量才会出现负值；而对于后者，即使是极其微弱的内生性问题 (如 $\theta=0.1$)，在 1000 次模拟中 h 统计量为负的比例仍然高达 16.3%。这与表 1 和表 2 中的分析结论一致，表明小样本偏误也是导致 h 统计量为负的一个原因，而内生性问题则会加剧小样本偏误的影响。同时，也是更为重要的，若按照本文的观点，即在 h 统计量呈现负值时拒绝原假设 (此时的检定力为 A 栏和 B 栏数据的加总)，则在上述两种面板结构下，只需分别满足 $\theta\geq 0.2$ 和 $\theta\geq 0.5$ 便可保证 h 统计量的检定力接近或大于 95%。

3) 一般意义下的内生性偏误下的模拟结果。前文讨论的内生性偏误 (B1) 是特指解释变量和个体效应相关，那么对于一般意义上的内生性偏误 (B2)，h 统计量的表现又如何呢？前文已经提到，在这种情况下，$\hat{\beta}_{FE}$ 和 $\hat{\beta}_{RE}$ 都是有偏的。由于此时的数据生成过程并未直接涉及 Hausman 检验的原假设 H_{RE}^2，因此可以推测此时 Hausman 检验无法很好地区分 FE 模型和 RE 模型之间的差异。表 4 中呈现的模拟结果证实了这一预期。

具体而言，对于 $(N,T)=(200,10)$ 和 $(1000,10)$ 这两种样本数较大的面板结构而言，无论 γ 取何值，h 统计量均未出现负值 (见 A 栏)。同时，h 统计量拒绝原假设的比例基本维持在 5% 左右，与我们设定的显著水平 5% 相当。而在 $(N,T)=(30,30)$ 面板结构中，仍然呈现出了前文反复提及的小样本偏误特征，在各种 γ 取值下，h 统计量出现负值的比例介于 4.8%–6.5% 之间，此时出现型一错误的概率也略微高于 5%，介于 7.8%–9.8% 之间。

因此，对于一般意义上的内生性偏误 (B2)，Hausman 检验是无法识别的。需要说的是，这并不意味着 h 统计量有问题，而是恰恰相反。它提示我们必须在满足 Hausman 检验的辅助假设条件 (H_{FE}^1 和 H_{RE}^1) 的前提下，才能应用 h 统计量进行统计推断，否则将得到虚假的结论，因为 H_{FE}^1 和 H_{RE}^1 并非 Hausman 检验的标的 [20:289]。

表 4　一般意义的内生性偏误下 Hausman 检验统计量的表现

γ 值	$N=200,T=10$		$N=1000,T=10$		$N=30,T=30$	
	A	B	A	B	A	B
0.1	0.00	6.00	0.00	5.20	4.80	7.80
0.2	0.00	4.50	0.00	4.90	6.20	9.10
0.3	0.00	6.10	0.00	5.50	6.10	7.90
0.4	0.00	4.80	0.00	4.70	5.20	8.60
0.5	0.00	5.40	0.00	4.60	5.50	9.80
0.6	0.00	7.30	0.00	4.90	5.80	9.80
0.7	0.00	6.40	0.00	6.20	5.40	9.50
0.8	0.00	5.30	0.00	4.70	6.50	8.10
0.9	0.00	5.20	0.00	4.30	5.90	9.40

注：A 代表检验统计量为负的比例，B 代表在 5% 显著水平下拒绝原假设的百分比。

4) 序列相关偏误下的模拟结果。当干扰项 ε_{it} 存在一阶序列相关时的模拟结果呈现于表 5。对于 $(N,T)=(200,10)$ 和 $(1000,10)$ 型面板结构而言，从 A 栏的结果可知，不管 ρ 取何值，h 统计量均未出现负值。而 B 栏中的结果则表明，h 统计量出现型一错误的概率与我们设定的显著水平 5% 非常接近。这意味着，虽然序列相关会在一定程度上影响式 (4) 中的方差协方差矩阵 $\hat{V}_{RE}$ 和 $\hat{V}_{FE}$ 的估计值，但并不会导致 h 统计量出现负值，也不会影响其检定力。当然，在 $(N,T)=(30,30)$ 型的面板结构下，h 统计量仍然表现出小样本偏误，与前文的分析结论相似，不再赘述。

表 5　序列相关偏误下 Hausman 检验统计量的表现

ρ 值	$N=200,T=10$		$N=1000,T=10$		$N=30,T=30$	
	A	B	A	B	A	B
0.1	0.00	6.10	0.00	5.10	4.90	7.60
0.2	0.00	4.10	0.00	5.30	6.30	8.40
0.3	0.00	6.40	0.00	5.00	6.10	7.50
0.4	0.00	5.00	0.00	4.70	5.50	8.90
0.5	0.00	5.00	0.00	4.50	5.90	9.00
0.6	0.00	6.50	0.00	4.90	5.20	11.00
0.7	0.00	5.80	0.00	6.10	5.20	10.00
0.8	0.00	6.00	0.00	4.60	6.20	8.40
0.9	0.00	6.80	0.00	3.70	8.70	8.80

注：A 代表检验统计量为负的比例，B 代表在 5% 显著水平下拒绝原假设的百分比。

表 6　非平行面板下 Hausman 检验统计量的表现

ω 值	$N=200,T=10$		$N=500,T=10$		$N=1000,T=10$	
	A	B	A	B	A	B
0.36	0.00	5.20	0.00	5.90	0.00	4.20
0.64	0.00	5.60	0.00	4.80	0.00	4.40
0.84	0.00	5.10	0.00	5.30	0.00	5.90
0.96	0.00	6.20	0.00	5.50	0.00	6.20

注：A 代表检验统计量为负的比例，B 代表在 5% 显著水平下拒绝原假设的百分比。

5) 非平行面板。表 6 呈现了非平行面板情形下的模拟结果。由于非平行面板主要出现在企业数据研究中，本文重点考察了三种“大 N 小 T”型的面板结构：$(N,\overline{T})=(200,10)$、$(500,10)$

和 (1000, 10)。借鉴 Bruno[27]，本文设定个四种 ω 值，以反映面板的非平行程度。可以看出，在所有面板结构设定下，无论其非平行程度较为严重 ($\omega=0.36$)，还是较为轻微 ($\omega=0.96$)，h 统计量均未出现负值，其出现型一错误的概率也接近 5%。这表明，面板非平行特征并不影响 h 统计量的小样本性质。

2.2 修正后的 Hausman 统计量的统计性质

由上述模拟分析可以初步推断，除了在所有设定下都普遍存在的小样本偏误外，导致 h 统计量出现负值的最主要的原因就是解释变量和个体效应相关 (违反假设 H_{RE})。此时，寻求修正方法就显得尤为必要。表 2 中的结果已经初步证实了修正后的统计量 h_{RE} 和 h_{FE} 在小样本下具有良好的表现。这里，我们进一步分析了二者在内生性偏误下的性质，结果分别呈现于表 7 中的 A 栏和 B 栏。

表 7 修正的 Hausman 检验统计量 h_{RE} 和 h_{FE} 的表现

θ 值	$N=200,T=10$		$N=1000,T=10$		$N=30,T=30$	
	A	B	A	B	A	B
A: h_{RE} 统计量						
0.1	0.00	13.80	0.00	48.80	0.00	8.50
0.2	0.00	38.10	0.00	97.30	0.00	21.00
0.3	0.00	74.70	0.00	100.00	0.00	46.40
0.4	0.00	95.40	0.00	100.00	0.00	69.80
0.5	0.00	99.60	0.00	100.00	0.00	86.60
0.6	0.00	100.00	0.00	100.00	0.00	97.60
0.7	0.00	100.00	0.00	100.00	0.00	99.80
0.8	0.00	100.00	0.00	100.00	0.00	100.00
0.9	0.00	100.00	0.00	100.00	0.00	100.00
B: h_{FE} 统计量						
0.1	0.00	13.80	0.00	48.80	0.00	8.60
0.2	0.00	38.20	0.00	97.30	0.00	21.00
0.3	0.00	74.70	0.00	100.00	0.00	46.50
0.4	0.00	95.40	0.00	100.00	0.00	70.00
0.5	0.00	99.60	0.00	100.00	0.00	86.60
0.6	0.00	100.00	0.00	100.00	0.00	97.60
0.7	0.00	100.00	0.00	100.00	0.00	99.80
0.8	0.00	100.00	0.00	100.00	0.00	100.00
0.9	0.00	100.00	0.00	100.00	0.00	100.00

注：A 代表检验统计量为负的比例，B 代表在 95% 置信水平下拒绝原假设的百分比。

可以看出，在所有设定下，h_{RE} 和 h_{FE} 统计量都没有出现小于 0 的异常情况。从检定力角度来看，在样本数较大的 $(N,T)=(1000,10)$ 型面板结构中，只要 $\theta\geq0.2$，h_{RE} 和 h_{FE} 统计量便能够非常灵敏地拒绝原假设，检定力大于 95%。即使在 $(N,T)=(200,10)$ 型面板结构中，当 $\theta\geq0.2$ 时，二者的检定力便会超过 95%。当然，在 $(N,T)=(200,10)$ 这种小样本设定下，h_{RE} 和 h_{FE} 统计量的检定力也会大幅降低。但相比于传统的 h 统计量，其表现已明显提高。整体而言，采用修正后的 h_{RE} 和 h_{FE} 统计量可以有效克服传统 h 统计量经常出现负值所带来的困扰，从而实现更为合理的统计推断。

Schreiber[8] 在理论分析基础上指出，当计算出的 h 统计量为负时，可以采用其绝对值，即

$|h|$ 进行统计推断。本文的模拟分析结果证实了这一理论预期。具体而言，对比表 7 和表 3 中结果，可以看出，在表 3 中，当 $\theta \geq 0.4$ 时，h 统计量出现负值的概率大幅增加，而从表 7 中的结果来看，当 $\theta = 0.4$ 时，我们可以在 95.4% 的概率下拒绝原假设。我们将这些小于零的 h 统计值取对数后进行假设检验，发现多数情况下都倾向于拒绝原假设。当然，诚如 Schreiber[8] 所言，这种方法只能在 h_{RE} 和 h_{FE} 统计量无法计算时作为一种次优选择。

3 扩展及建议

为克服 h 统计量为负这一局限，文献中还提出了多种替代检验方法 [19,29–30]。其中，应用较为广泛的是 Arellano[19] 基于过度识别检验提出的 Wald 统计量 (简称为 h_W)。本小节重点探讨 h_W 统计量的小样本性质。

h_W 统计量的构造原理和估算方法都很简单。如前所述，FE 和 RE 估计量都需要满足一般意义上的外生性假设条件 H_{FE}^1 和 H_{RE}^1，即 $E(\varepsilon_{it}|\mathbf{x}_i, \mu_i) = 0$，而 RE 还需要进一步满足面板特定的外生性假设条件 H_{RE}^2，即 $E(\mu_i|\mathbf{x}_i) = E(\mu_i) = 0$。显然，这个新增的正交条件 (orthogonality conditions) 可以视为一个过度识别约束 (overidentifying restrictions)，它可以帮助我们区分 RE 的前提假设是否合理。我们可以估计如下模型来构造 h_W 统计量：

$$\begin{pmatrix} \tilde{y}_{it} \\ \check{y}_{it} \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{x}}_{it} & 0 \\ \check{\mathbf{x}}_{it} & \check{\mathbf{x}}_{it} \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \begin{pmatrix} \tilde{v}_{it} \\ \check{v}_{it} \end{pmatrix}, \tag{13}$$

其中，$\check{y}_{it} = (y_{it} - \overline{y}_i)$，$\tilde{y}_{it} = (y_{it} - \hat{\theta}_i \overline{y}_i)$，$\tilde{\mathbf{x}}_{it}$ 和 $\check{\mathbf{x}}_{it}$ 具有相似的定义方式。需要说明的是，本文所描述的 h_W 统计量与 Arellano[19] 略有差异。我们是根据 FE 和 RE 估计量来构造 h_W 统计量的，而 Arellano[19] 则是采用 FE 和组间估计量 (between group estimator) 构造的。当然，这并不影响最终的统计推断。显然，在 (13) 式中，β 的 OLS 估计即为 RE 估计量 $\hat{\beta}_{RE}$，而 γ 的 OLS 估计则是 $\hat{\beta}_{RE}$ 和 $\hat{\beta}_{FE}$ 之间的差异：

$$\hat{\gamma} = \hat{\beta}_{FE} - \hat{\beta}_{RE}. \tag{14}$$

因此，对约束条件 $\gamma = 0$ 执行 Wald 检验，即可得到 h_W 统计量。在同方差假设下，其渐进分布特征与 h 统计量相同。对于平行面板而言，二者具有相同的数值。显然，由于构造过程非常简单，h_W 统计量可以很容易扩展到考虑异方差、序列相关或截面相关等稳健型估计量的情形 (详见 Arellano[19]、Wooldridge[20:290–291])，这是它相对于传统 h 统计量的一个优势。当然，h_W 统计量的最大优势是不会出现负值。

前述分析表明，内生性偏误 (B1) 是导致 Hausman 检验失效 (h 统计量出现负值) 的主要原因。那么，在这种情况下，h_W 统计量的表现如何呢？相应的 Monte Carlo 模拟结果呈现于表 8。可以看出，此处的模拟结果与表 7 第 B 栏非常相似。在所有设定下，h_W 统计量都没有出现小于 0 的情况。同时，即使在样本量适中 (如 $N \times T = 200 \times 10$) 的情况下，其检定力仍然较高。这意味着，$\mathrm{h}_W$ 统计量可以作为传统 Hausman 检验的一个很好的替代方法。我们还进一步在偏误 B2–B4，以及异方差设定下模拟了 h_W 统计量的小样本性质，均未出现小于零的情形。而且，在样本量较大 (如 $N \times T = 1000 \times 10$) 的情况下，其检定力都较高。受限于篇幅，相关的结果未能呈现。

表 8 内生性偏误下 h_W 统计量的表现

θ 值	$N=200,T=10$		$N=1000,T=10$		$N=30,T=30$	
	A	B	A	B	A	B
0.1	0.00	13.80	0.00	48.80	0.00	8.60
0.2	0.00	38.20	0.00	97.30	0.00	21.00
0.3	0.00	74.70	0.00	100.00	0.00	46.50
0.4	0.00	95.40	0.00	100.00	0.00	70.00
0.5	0.00	99.60	0.00	100.00	0.00	86.60
0.6	0.00	100.00	0.00	100.00	0.00	97.60
0.7	0.00	100.00	0.00	100.00	0.00	99.80
0.8	0.00	100.00	0.00	100.00	0.00	100.00
0.9	0.00	100.00	0.00	100.00	0.00	100.00

注：A 代表检验统计量为负的比例，B 代表在 95% 置信水平下拒绝原假设的百分比。

4 结论

本文采用 Monte Carlo 模拟研究了 Hausman 检验统计量的小样本性质。研究结果表明，内生性问题 (即，解释变量 x 与个体效应 μ 相关) 是导致传统 h 统计量出现负值的主要原因。这意味着，Hausman 检验出现负值可以视为拒绝原假设的信号。这与一些主流的教科书以及计量软件 (如 Stata) 所持的观点恰好相反。

本文的主要贡献可以归结为如下两个方面：一方面，本文通过 Monte Carlo 模拟分析，为 Schreiber[8]、Magazzini 和 Calzolari[9] 等学者的理论分析提供了基于小样本的经验支持。虽然本文仅针对面板数据模型中的固定效应和随机效应模型研究了 h 统计量的小样本性质，但本文的方法可以很方便地扩展到其他模型中，如 Logit 模型、处理效应模型等。另一方面，本文的研究结果为大量应用面板数据模型进行实证研究的文献提供了模型筛选的判断依据，这在前期文献中是一个令学者们极度困扰的问题。

对于实证分析而言，从本文研究中能得到如下启示。其一，我们发现，在小样本下 ($N\times T=30\times 30$)，本文所考察的所有统计量的检定力都非常低，这意味着，在使用中国省级面板数据进行 Hausman 检验时，要非常慎重地进行统计推断，保守的处理方法是采用 FE 估计量。其二，由于 Hausman 检验失效的主要原因是内生性问题，因此，当出现负值时，应该重点考虑模型的设定问题。例如，很多变量都具有动态特征，如果将动态模型错误地设定为静态模型，就会直接导致内生性偏误。推而广之，遗漏变量、衡量偏误都可能导致内生性问题。

在 Hausman 检验失效的情况下，有多种替代检验方法。一是采用 Cameron 和 Trivedi[22:260] 建议的修正 Hausman 统计量 (h_{RE} 或 h_{FE})，该统计量不会出现负值，在样本数不至于过少的情况下具有较高的检定力。另一种是 Arellano[19] 提出的 Wald 统计量 (h_W)。该统计量同样可以避免出现负值的情形。同时，它非常容易计算，并能很方便地扩展到异方差、序列相关、截面相关等情形，既可以视为传统 Hausman 检验的一个替代方法，也可以视为对其的一个重要扩展。

[参考文献]

[1] Hausman J. Specification tests in econometrics [J]. Econometrica, 1978, 46(6): 1251–1271.

[2] 王敏, 梁利. 中国农民消费行为及影响因素分析 [J]. 数理统计与管理, 2010, (5): 780–788.

[3] 柯孔林. 中国银行业市场竞争结构测度：基于 bresnahan 范式研究 [J]. 数理统计与管理, 2010, 29(4): 678–686.

[4] Cameron A, Trivedi P. Microeconometrics: Methods and Applications [M]. Cambridge University Press, 2005.

[5] Baltagi B. Econometric Analysis of Panel Data [M]. Chichester: John Wiley & Sons, 2008.

[6] Hsiao C. Analysis of Panel Data [M]. Cambridge University Press, 2003.

[7] StataCorp. Stata Release 11: Stata base reference manual [J]. College Station, TX: StataCorp LP, 2009: 1239–1260.

[8] Schreiber S. The Hausman test statistic can be negative even asymptotically [J]. Journal of Economics and Statistics, 2008, 228(4): 394–405.

[9] Magazzini L, Calzolari G. Negative variance estimates in panel data models [D]. Working papers, Department of Economics University of Verona, 2010.

[10] 金煜, 陈钊, 陆铭. 中国的地区工业集聚: 经济地理, 新经济地理与经济政策 [J]. 经济研究, 2006, (4): 79–89.

[11] 蒋文广. 外国直接投资对中国技术创新的影响 — 基于地区层面的研究 [J]. 财经研究, 2007, (6): 4–17.

[12] 胡向婷, 张璐. 地方保护主义对地区产业结构的影响 [J]. 经济研究, 2005, (2): 102–112.

[13] 沈根祥. 沪深 300 指数日内跳的 hausman 检验 [J]. 数理统计与管理, 2010, 29(4): 713–718.

[14] 张卉, 詹宇波, 周凯. 集聚、多样性和地区经济增长：来自中国制造业的实证研究 [J]. 世界经济文汇, 2007, (3): 16–29.

[15] 李婉丽, 吕怀立. 股权制衡与上市公司派现水平的联动关系研究 [J]. 管理学报, 2009, (1): 84–90.

[16] 孙文杰, 沈坤荣. 技术引进与中国企业的自主创新：基于分位数回归模型的经验研究 [J]. 世界经济, 2007, (11): 32–43.

[17] 赵国庆, 任宇宁. 基于面板数据的中日电力行业规模收益比较分析 [J]. 商业经济与管理, 2009, (3): 61–67.

[18] Hayashi F. Econometrics [M]. Princeton University Press, 2000.

[19] Arellano M. On the testing of correlated effects with panel data [J]. Journal of Econometrics, 1993, 59(1–2): 87–97.

[20] Wooldridge J. Econometric Analysis of Cross Section and Panel Data [M]. MIT press, 2002.

[21] StataCorp. Stata longitudinal-data/panel-data reference manual [J]. College Station, TX: StataCorp LP, 2009, Release 11: 464–465.

[22] Cameron A, Trivedi P. Microeconometrics Using Stata [M]. Stata Press, 2009.

[23] Baltagi B, Bresson G, Pirotte A. Fixed effects, random effects or hausman-taylor? A pretest estimator [J]. Economics Letters, 2003, 79(3): 361–369.

[24] Baltagi B, Song S. Unbalanced panel data: A survey [J]. Statistical Papers, 2006, 47(4): 493–523.

[25] Heckman J. Sample selection bias as a specification error [J]. Econometrica, 1979, 47(1): 153–161.

[26] Baltagi B H, Chang Y J. Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model [J]. Journal of Econometrics, 1994, 62(2): 67–89.

[27] Bruno G S F. Approximating the bias of the Lsdv estimator for dynamic unbalanced panel data models [J]. Economics Letters, 2005, 87(3): 361–366.

[28] Ahrens H, Pincus R. On two measures of unbalancedness in a one-way model and their relation to efficiency [J]. Biometrical Journal, 1981, 23(3): 227–235.

[29] Hausman J, Taylor W. Panel data and unobservable individual effects [J]. Econometrica, 1981, 49(6): 1377–1398.

[30] Chamberlain G. Multivariate regression models for panel data [J]. Journal of Econometrics, 1982, 18(1): 5–46.