



CENTRE FOR **STOCHASTIC GEOMETRY**  
AND ADVANCED **BIOIMAGING**



Adrian Baddeley, Rolf Turner and Ege Rubak

**Adjusted composite likelihood ratio test  
for spatial Gibbs point processes**

No. 13, November 2014

# Adjusted composite likelihood ratio test for spatial Gibbs point processes

Adrian Baddeley<sup>1</sup>, Rolf Turner<sup>2</sup> and Ege Rubak<sup>3</sup>

<sup>1</sup>Centre for Exploration Targeting, University of Western Australia,  
adrian.baddeley@uwa.edu.au

<sup>2</sup>Department of Statistics, University of Auckland, r.turner@auckland.ac.nz

<sup>3</sup>Department of Mathematical Sciences, University of Aalborg, rubak@math.aau.dk

## Abstract

We investigate an analogue of the likelihood ratio test for spatial Gibbs point process models fitted by maximum pseudolikelihood or maximum composite likelihood. The test statistic must be adjusted in order to obtain an asymptotic  $\chi^2$  distribution under the null hypothesis. Adjustments developed for composite likelihoods of finite systems of random variables are adapted to the point process setting. Recent results in point process theory are used to estimate the composite information  $J$  and sensitivity  $H$  from data. In a large simulation experiment we find that the proposed test is exact if  $J$  and  $H$  are known exactly; it is slightly conservative when  $J$  and  $H$  are estimated from data.

*Keywords:* Georgii-Nguyen-Zessin formula; Godambe-Heyde criterion; Moment matching; Papangelou conditional intensity; Pseudolikelihood; Score test statistic; Variance estimation.

## 1 Introduction

Statistical inference for spatial point processes [1, 2, 3] is difficult because, apart from the Poisson point process, most of the useful models have intractable likelihoods [4, 5]. It is therefore common to fit point process models to data by maximising a *composite likelihood* [6], such as Besag's [7] pseudolikelihood for point processes with tractable conditional intensities, or the Palm likelihood [8, 9, 10] or second order composite likelihood [11] for models with tractable pair correlation functions. This is a good practical strategy: the computations are extremely fast [12], the parameter estimates are the solutions of unbiased estimating equations, and they are asymptotically normal under suitable conditions [13, 14].

However, formal inference such as model selection and hypothesis testing is problematic using these composite likelihoods, because the analogue of the likelihood ratio test statistic based on composite likelihoods does not have an asymptotic  $\chi^2$  distribution under the null hypothesis. Rather, the asymptotic null distribution is that of a weighted sum of independent  $\chi^2$  variables.

The literature on composite likelihood for *finite* systems of random variables contains several proposals for adjusting the log composite likelihood ratio so as to obtain a null distribution which is approximately  $\chi^2$ . Pace *et al* [15] review these proposals and offer a new one which performs very well in their examples.

Our aim is to apply these adjustments to the composite likelihoods of spatial point process models of Gibbs type with tractable conditional intensity. In principle this should be possible – even though a spatial point process is a *continuum* of random variables – because the adjustments depend mainly on the asymptotic normality of the composite score, which typically holds in the point process context as well.

In practice, the adjustments discussed in [15] are not directly applicable to Gibbs point processes, because the composite information  $J$  (variance of the composite score) and the Hessian  $H$  (negative expectation of the derivative of the composite score) are not known explicitly as functions of the parameter  $\theta$ .

For Gibbs point processes, estimators of  $H$  and  $J$  have recently been developed [16, 17, 18], so it is now possible to perform “plug-in” versions of the composite likelihood adjustments, in which the true  $H$  and  $J$  are approximated by data-dependent estimators. It is an open question whether plug-in versions of the adjustments will be satisfactory.

This paper gives explicit formulae for the adjustments for the pseudolikelihood ratio test for Gibbs point processes, and reports the result of a large simulation experiment assessing the performance of the adjustments. We find that the plug-in versions of the adjustments perform reasonably well except for a “breakdown” occurring when interaction between points is very strong.

Section 2 summarises adjustments to the composite likelihood ratio test statistic that are known in the literature. Section 3 describes composite likelihoods for spatial Gibbs point processes including recent results about variance. Section 4 spells out the proposed procedure for adjusting the composite likelihood ratio test for a spatial Gibbs point process. Section 5 reports on the results of a large experiment. We end with a discussion.

## 2 Adjustments to composite likelihood ratio test

The general context is one in which a parametric model governed by a  $d$ -dimensional parameter  $\theta$  is fitted to data  $\mathbf{x}$  by maximising a pseudolikelihood or composite likelihood  $\text{CL}(\theta, \mathbf{x})$ . We assume there are no constraints on  $\theta$  so the maximum is found as the root of the composite score  $U(\theta, \mathbf{x}) = (\partial/\partial\theta) \log \text{CL}(\theta, \mathbf{x})$ . The variance/covariance matrix of the composite score  $J(\theta) = \text{var}[U(\theta, \mathbf{X})]$  is the ‘composite information’, and the matrix  $H$

$$H(\theta) = \mathbb{E}_\theta \left[ -\frac{\partial}{\partial\theta} U(\theta, \mathbf{X}) \right] \quad (2.1)$$

is the ‘sensitivity’. It is desired to test the null hypothesis  $\mathcal{H}_0 : \theta \in \Theta_0$  against the alternative  $\mathcal{H}_1 : \theta \in \Theta_1$  where  $\Theta_0 \subset \Theta_1$ . A natural choice for the test statistic would

be the analogue of the likelihood ratio test statistic

$$\Lambda = 2 \log \frac{\text{CL}(\hat{\theta}, \mathbf{x})}{\text{CL}(\check{\theta}, \mathbf{x})} \quad (2.2)$$

where  $\hat{\theta} = \arg \max_{\theta \in \Theta_1} \text{CL}(\theta, \mathbf{x})$  and  $\check{\theta} = \arg \max_{\theta \in \Theta_0} \text{CL}(\theta, \mathbf{x})$  are the maximum composite likelihood estimates of  $\theta$  under the alternative and null hypotheses respectively. However,  $\Lambda$  does not have a asymptotic  $\chi^2$  distribution under  $\mathcal{H}_0$  as it would in the case of the likelihood.

For composite likelihoods of finite systems of random variables, Pace et al [15, eq. (3.8), (3.9)] proposed the following adjustments to  $\Lambda$  which are designed to bring it closer to having a  $\chi^2$  distribution.

## 2.1 Simple null hypothesis

In the case of a simple null hypothesis  $\mathcal{H}_0 : \theta = \theta_0$ , the asymptotic distribution of  $\Lambda$  is that of  $\sum_j \lambda_j \chi_1^2$  where  $\lambda_j$ ,  $j = 1, \dots, d$  are the (not necessarily distinct) eigenvalues of  $J(\theta_0)H(\theta_0)^{-1}$ . In particular the mean of this asymptotic distribution is  $m = \sum_j \lambda_j = \text{trace}(J(\theta_0)H(\theta_0)^{-1})$  [19, Thm 3.1]. The *first moment matching* adjustment [20], [21, sec. 9.3.3] is

$$\Lambda^* = \frac{d}{m} \Lambda. \quad (2.3)$$

This simply rescales  $\Lambda$  so that its mean is equal to  $d$ , the mean of  $\chi_d^2$ . For brevity we call this the ‘‘mean adjustment’’.

The adjustment proposed by Pace *et al* is [15, eq. (3.8)]

$$\Lambda^* = \frac{U(\theta_0, \mathbf{x})^\top J(\theta_0)^{-1} U(\theta_0, \mathbf{x})}{U(\theta_0, \mathbf{x})^\top H(\theta_0)^{-1} U(\theta_0, \mathbf{x})} \cdot \Lambda \quad (2.4)$$

which we will call the ‘‘Pace-Salvan-Sartori adjustment’’ or ‘‘PSS adjustment’’. Notice that the numerator  $U(\theta_0, \mathbf{x})^\top J(\theta_0)^{-1} U(\theta_0, \mathbf{x})$  is the composite score test statistic, which is asymptotically  $\chi^2$  with  $d$  degrees of freedom under  $\mathcal{H}_0$  [21, p. 193], while the denominator  $U(\theta_0, \mathbf{x})^\top H(\theta_0)^{-1} U(\theta_0, \mathbf{x})$  is a first order approximation to  $\Lambda$ , so this adjustment effectively matches the first order asymptotic behaviour of  $\Lambda$  to that of the score test statistic.

## 2.2 Composite null hypothesis

Suppose that  $\theta = (\varphi, \psi)$  where  $\psi$  is a  $d_0$ -dimensional parameter of interest and consider the composite null hypothesis  $\mathcal{H}_0 : \psi = \psi_0$ . Let the matrices  $H(\theta)$  and  $J(\theta)$  be as defined above, and define the Godambe-Heyde criterion matrix

$$G(\theta) = H(\theta)J(\theta)^{-1}H(\theta).$$

Let  $U_\psi = (\partial/\partial\psi) \log \text{CL}(\theta)$  be the component of the score corresponding to  $\psi$ . Following [15] we write  $H^{\psi\psi}(\theta)$  for the  $(\psi, \psi)$  submatrix of  $H(\theta)^{-1}$ , the *inverse* of the

sensitivity matrix, and  $G^{\psi\psi}(\theta)$  for the  $(\psi, \psi)$  submatrix of  $G(\theta)^{-1}$ , the *inverse* of the Godambe-Heyde criterion matrix.

The asymptotic null distribution of  $\Lambda$  is that of  $\sum_j \lambda_j \chi_1^2$  where  $\lambda_j$ ,  $j = 1, \dots, d_0$  are the (not necessarily distinct) eigenvalues of  $(H^{\psi\psi})^{-1}G^{\psi\psi}$ . The mean adjustment is then

$$\Lambda^* = \frac{d_0}{m} \Lambda \quad (2.5)$$

where  $m = \sum_j \lambda_j = \text{trace}((H^{\psi\psi})^{-1}G^{\psi\psi})$ . The score test statistic is [15, eq. (2.4)]

$$S = U_\psi(\check{\theta})^\top H^{\psi\psi}(\check{\theta}) G^{\psi\psi}(\check{\theta})^{-1} H^{\psi\psi}(\check{\theta}) U_\psi(\check{\theta}) \quad (2.6)$$

where  $\check{\theta} = (\check{\varphi}(\psi_0), \psi_0)$  is the maximum composite likelihood estimate of  $\theta$  under the null hypothesis. Then the PSS adjustment is [15, eq. (3.9)]

$$\Lambda^* = \frac{S}{U_\psi(\check{\theta})^\top H^{\psi\psi}(\check{\theta}) U_\psi(\check{\theta})} \Lambda. \quad (2.7)$$

The corrected statistic  $\Lambda^*$  has asymptotic null distribution  $\chi_{d_0}^2$  approximately. When  $\psi$  is one-dimensional, the mean adjustment (2.5) and the PSS adjustment (2.7) both collapse to the same quantity [15, p. 135]

$$\Lambda^* = \frac{H^{\psi\psi}(\check{\theta})}{G^{\psi\psi}(\check{\theta})} \Lambda \quad (2.8)$$

which has asymptotic distribution *exactly*  $\chi_1^2$ .

To avoid confusion we note that, because  $H^{\psi\psi}$  and  $G^{\psi\psi}$  are submatrices of  $H^{-1}$  and  $G^{-1}$  respectively, (2.8) has the same general form as (2.4).

### 3 Likelihood devices for Gibbs point processes

In the point process setting [5, 3, 2] the dataset  $\mathbf{x} = \{x_1, \dots, x_n\}$  consists of points in a bounded subset (window)  $W$  of  $\mathbb{R}^m$ , and  $\mathbf{x}$  is assumed to be a realisation of a Gibbs point process  $\mathbf{X}$ . In applications  $m$  is typically 2 or 3 (and all the experiments in Section 5 have  $m = 2$ ), but the theory holds for any natural number  $m$ . The distribution of the process is determined by its Papangelou conditional intensity  $\lambda(u, \mathbf{x})$  as explained in [5, 3]. For simplicity we consider a model in which the conditional intensity is loglinear in the parameter,

$$\lambda_\theta(u, \mathbf{x}) = \exp(\theta^\top T(u, \mathbf{x})). \quad (3.1)$$

#### 3.1 Pseudolikelihood

Besag [7] defined the *pseudolikelihood* of the finite point process with Papangelou conditional intensity  $\lambda_\theta(u, \mathbf{x})$  as

$$\text{PL}(\theta, \mathbf{x}) = \left[ \prod_{i=1}^n \lambda_\theta(x_i, \mathbf{x}) \right] \exp\left(-\int_W \lambda_\theta(u, \mathbf{x}) \, du\right). \quad (3.2)$$

Fast, practical algorithms exist for fitting models by maximising the pseudolikelihood [22, 12]. For the loglinear model (3.1) the pseudolikelihood score is

$$U(\theta, \mathbf{x}) = \frac{\partial \text{PL}}{\partial \theta} = \sum_{i=1}^n T(x_i, \mathbf{x}) - \int_W T(u, \mathbf{x}) \lambda_\theta(u, \mathbf{x}) du. \quad (3.3)$$

The pseudolikelihood score is an unbiased estimating function, by virtue of the Georgii-Nguyen-Zessin formula [23, 24]. Under reasonable conditions, the pseudolikelihood score evaluated at the true value  $\theta$  is asymptotically normal, and the maximum pseudolikelihood estimator of  $\theta$  is consistent and asymptotically normal [7, 13, 14].

The sensitivity (2.1) is

$$H(\theta) = \mathbb{E}_\theta \int_W T(u, \mathbf{X}) T(u, \mathbf{X})^\top \lambda_\theta(u, \mathbf{X}) du = \mathbb{E}_\theta \sum_{i=1}^n T(x_i, \mathbf{X}) T(x_i, \mathbf{X})^\top. \quad (3.4)$$

Two estimators of  $H(\theta)$  are

$$\hat{H}(\theta) = \int_W T(u, \mathbf{X}) T(u, \mathbf{X}) T(u, \mathbf{X}) T(u, \mathbf{X})^\top \lambda_\theta(u, \mathbf{X}) du \quad (3.5)$$

$$\text{and } \tilde{H} = \sum_{i=1}^n T(x_i, \mathbf{X}) T(x_i, \mathbf{X})^\top. \quad (3.6)$$

Under a simple null hypothesis  $H_0 : \theta = \theta_0$ , both  $\hat{H}(\theta_0)$  and  $\tilde{H}$  are unbiased for  $H(\theta_0)$ . Under a composite null  $H_0 : \theta \in \Theta_0$ ,  $\hat{H}(\hat{\theta})$  is *approximately* unbiased for  $H(\hat{\theta})$  and  $\tilde{H}$  is unbiased for  $H(\theta)$ , where  $\theta$  is the true value of the parameter governing the distribution of  $\mathbf{X}$ .

Formulae for the variance of  $U$ , and estimators of this variance, were derived recently [16, 17]. The variance of the pseudolikelihood score (3.3) for fixed  $\theta$  is

$$\text{var}[U(\theta, \mathbf{X})] = A_1 + A_2 + A_3 \quad (3.7)$$

where  $A_1 = H(\theta)$  and

$$A_2 = \mathbb{E}_\theta \int_W \int_W T(u | \mathbf{X}) T(v | \mathbf{X})^\top \left( \lambda_\theta(u | \mathbf{X}) \lambda_\theta(v | \mathbf{X}) - \lambda_\theta^{[2]}(u, v | \mathbf{X}) \right) du dv \quad (3.8)$$

$$A_3 = \mathbb{E}_\theta \int_W \int_W \Delta_v T(u | \mathbf{X}) \Delta_u T(v | \mathbf{X})^\top \lambda_\theta^{[2]}(u, v | \mathbf{X}) du dv \quad (3.9)$$

with the second order Papangelou conditional intensity given by

$$\lambda_\theta^{[2]}(u, v | \mathbf{X}) = \lambda_\theta(u | \mathbf{X} \setminus \{v\}) \lambda_\theta(v | \mathbf{X} \cup \{u\}),$$

and where  $\Delta$  is the increment operator for set functions,

$$\Delta_u T(v | \mathbf{X}) = T(v | \mathbf{X} \cup \{u\}) - T(v | \mathbf{X}).$$

Consistent estimators of  $\text{var}[U(\theta, \mathbf{X})]$  can be constructed by choosing consistent estimators of the quantities  $A_1, A_2, A_3$  in (3.7) based on the identities (3.4) and (3.8)–(3.9) as described in [17]. The estimators used in our experiments are (3.5), (3.6) and

$$\hat{A}_2 = \sum_i \sum_{j \neq i} T(x_i | \mathbf{x} \setminus x_j) T(x_j | \mathbf{x} \setminus x_i)^\top \left( \frac{\lambda_\theta(x_j | \mathbf{x} \setminus x_i)}{\lambda_\theta(x_j | \mathbf{x})} - 1 \right) \quad (3.10)$$

$$\hat{A}_3 = \sum_i \sum_{j \neq i} \Delta_{x_i} T(x_j | \mathbf{x} \setminus x_i) \Delta_{x_j} T(x_i | \mathbf{x} \setminus x_j)^\top. \quad (3.11)$$

These four estimators are unbiased and consistent if the model is true (with the correct value of  $\theta$ ) while  $\hat{H}(\theta)$  and  $\hat{A}_3$  are unbiased and consistent even under a misspecified model, by the Campbell-Mecke formula [25], [26, (4.4.3)].

### 3.2 Logistic likelihood

Baddeley *et al* [18] (see also [27]) proposed the *logistic likelihood*

$$\text{LOL}(\theta, \mathbf{x}) = \left[ \prod_{i=1}^n \frac{\lambda_\theta(x_i, \mathbf{x})}{\lambda_\theta(x_i, \mathbf{x}) + \rho(x_i)} \right] \left[ \prod_{j=1}^m \frac{\rho(y_j)}{\lambda_\theta(y_j, \mathbf{x}) + \rho(y_j)} \right] \quad (3.12)$$

where  $\mathbf{y} = \{y_1, \dots, y_m\}$  is a randomly-generated set of “dummy points”, a realisation of a point process  $\mathbf{Y}$  with intensity function  $\rho$ . They showed that the composite score associated with (3.12) is an unbiased estimating function, and that under reasonable conditions, the composite score evaluated at the true  $\theta$  is asymptotically normal, and the maximum composite likelihood estimator is consistent and asymptotically normal.

Expressions for the composite information  $J$  and sensitivity  $H$  for the logistic likelihood, and consistent estimators of these quantities, are given in [18].

## 4 Adjusted composite likelihood ratio test statistic for Gibbs point process

Here we spell out the procedure for adjusting the composite likelihood ratio test statistic for a point process model, in the case of the pseudolikelihood (3.2). Assume a Gibbs point process model with loglinear conditional intensity (3.1).

We fit the null and alternative models by maximising the pseudolikelihood (3.2), and evaluate the uncorrected test statistic  $\Lambda$  in (2.2) where CL is PL.

### 4.1 Simple null

In the case of a simple null hypothesis  $\mathcal{H}_0 : \theta = \theta_0$ , we compute the pseudolikelihood score  $U(\theta_0, \mathbf{x})$  of the alternative model at the null value, estimate the sensitivity  $H(\theta_0, \mathbf{x})$  of the alternative model at the null value using either (3.5) or (3.6), and estimate the composite information  $J(\theta_0) = \text{var}[U(\theta_0, \mathbf{X})]$ , the variance of the score

of the alternative model at the null value, by plugging into (3.7) the estimates  $\widehat{A}_1, \widehat{A}_2, \widehat{A}_3$  given by (3.10)–(3.11) and either (3.5) or (3.6). Then we compute the mean adjustment (2.3) where  $d$  is the dimension of  $\theta$  and  $m = \text{trace}(\widehat{J}\widehat{H}^{-1})$ , and the PSS adjustment (2.4).

## 4.2 Composite null

In the case of a composite null hypothesis  $\mathcal{H}_0 : \psi = \psi_0$  where  $\theta = (\varphi, \psi)$ , calculations are performed for the fitted null model, with parameter  $\check{\theta} = (\check{\varphi}, \psi_0)$  obtained by maximising the pseudolikelihood over the nuisance parameter  $\varphi$ . We compute the pseudolikelihood score  $U(\check{\theta}, \mathbf{x})$  of the alternative model at the null MPLE, and estimate  $H(\check{\theta}, \mathbf{x})$  and  $J(\check{\theta}) = \text{var}[U(\check{\theta}, \mathbf{X})]$  by  $\widehat{H}, \widehat{J}$  using the same expressions (3.5) or (3.6) and (3.10)–(3.11). We form the Godambe-Heyde information  $\widehat{G} = \widehat{H} \widehat{J}^{-1} \widehat{H}$ , and extract  $\widehat{H}^{\psi\psi}, \widehat{G}^{\psi\psi}$  which are the submatrices of the inverses  $\widehat{H}^{-1}$  and  $\widehat{G}^{-1}$  corresponding to the parameter of interest  $\psi$ . Then we compute the mean adjustment (2.5) where  $d_0$  is the dimension of  $\psi$ , and  $m = \text{trace}((\widehat{H}^{\psi\psi})^{-1} \widehat{G}^{\psi\psi})$ , and the PSS adjustment (2.7).

## 5 Experimental results

We performed simulation experiments to evaluate the true null distributions of the adjusted statistics.

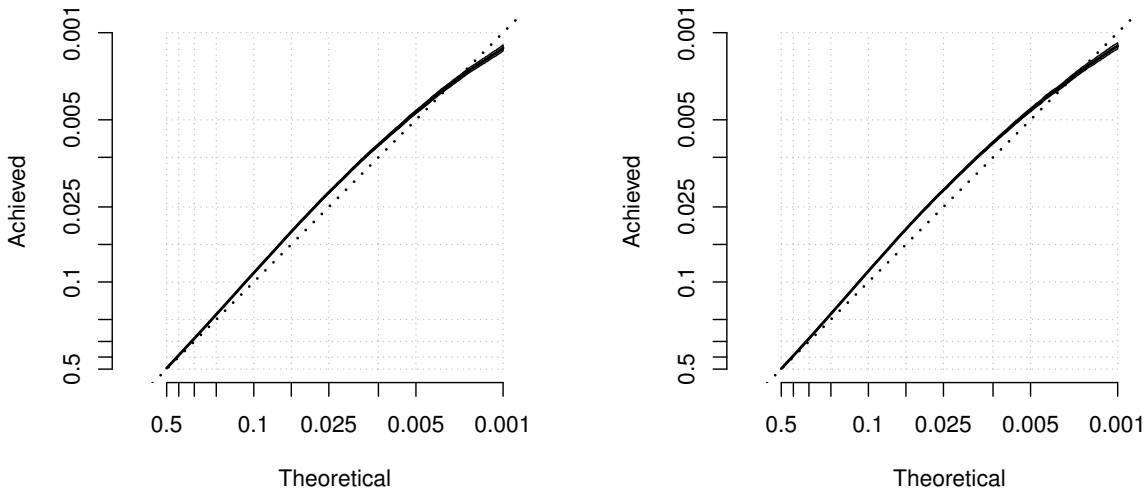
### 5.1 Simple null hypothesis

The null hypothesis was taken to be a Strauss point process [28, 29, 5] in the unit square. The interaction radius  $r = 0.05$  was fixed and was not treated as a parameter. The parameters of the Strauss model are then  $\theta = (\beta, \gamma)$  where the interaction strength  $\gamma$  penalises pairs of points within distance  $r$  of each other by a factor  $\gamma$ . Thus when  $\gamma$  is close to zero almost no pairs of points will be within distance  $r$  of each other while the points will appear completely at random when  $\gamma = 1$  (values above one yield an invalid model). The first order parameter  $\beta$  mainly adjusts the intensity of points (but the parameter is not directly interpretable as an intensity). In our experiment we used the parameter values  $\theta_0 = (\beta_0, \gamma_0)$  where  $\gamma_0$  ranged from 0.05 to 0.95 in steps of 0.05, then from 0.95 to 1 in steps of 0.01, and for each value of  $\gamma_0$ , we determined  $\beta_0$  so that the intensity (approximated by the Poisson-saddlepoint approximation [30, 31]) was equal to 100.

For each value of  $\theta_0$ , we generated 1 million realisations of the null model using a perfect simulation algorithm [32] on a supercomputing cluster. The model was fitted to each simulated realisation by maximum pseudolikelihood [22] using the correct value of  $r$ , and we evaluated the unadjusted pseudolikelihood ratio test statistic  $\Lambda$ , the null score  $U(\theta_0; \mathbf{x})$ , the estimated Hessian  $\widehat{H}(\theta_0; \mathbf{x})$  from (3.5), and the estimate  $\widehat{J}(\theta_0)$  of the composite information from (3.5) and (3.10)–(3.11). From these we calculated the adjusted composite likelihood ratio test statistics  $\Lambda^*$  of (2.3) and (2.4) for each simulation, plugging-in the estimates of  $H$  and  $J$ . Referring these



adjusted statistics to the  $\chi^2$  distribution with  $d = 2$  degrees of freedom, we obtained the  $p$ -values  $p^* = \mathbb{P}\{\chi_2^2 > \Lambda^*\}$ .



**Figure 1:** Logarithmic plot of actual significance against nominal significance for the  $\chi^2$  test of the simple null hypothesis  $H_0 : \gamma = 0.5$  based on the adjusted composite likelihood ratio test statistic, with matrices  $H$  and  $J$  estimated from data. *Left:* first moment matching adjustment. *Right:* Pace-Salvan-Sartori adjustment. Thickness of curve indicates 95% confidence interval for true significance.

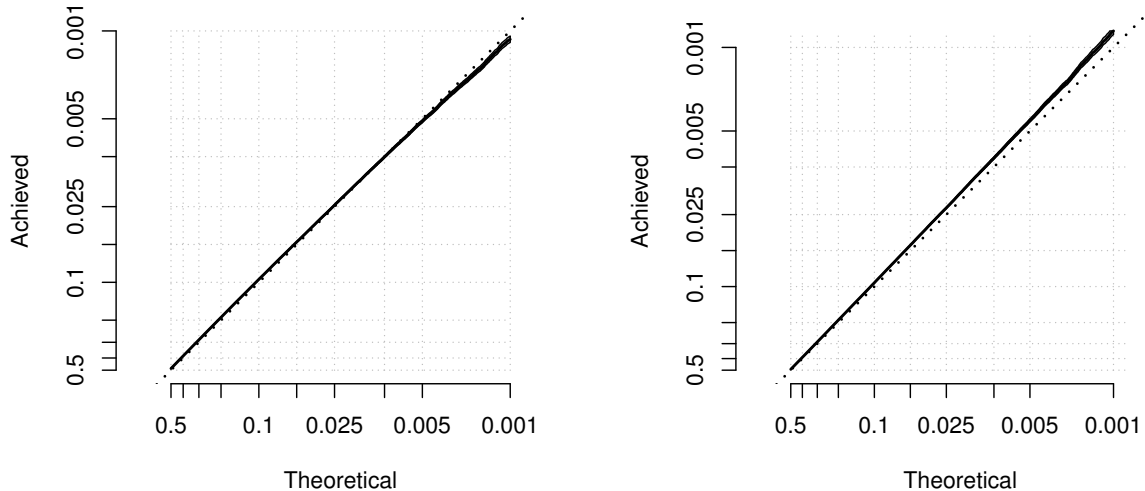
Figure 1 compares the empirical cumulative distribution function of  $p^*$  (based on 1 million values) against the uniform distribution, on a logarithmic scale, for the case  $\gamma = 0.5$ . This is effectively a plot of actual significance against nominal significance for the  $\chi^2$  test. The thickness of the solid curve indicates the 95% confidence interval for the true significance level (based on the asymptotic normal distribution of the proportions in this large experiment).

In order to check that the adjustments would work correctly for point process pseudolikelihood ratios if the matrices  $J(\theta_0)$  and  $H(\theta_0)$  were known exactly, we first computed very accurate approximations to  $J(\theta_0)$  and  $H(\theta_0)$  by averaging the 1 million estimates obtained from the simulations. These were then used to calculate the adjusted composite likelihood ratio test statistics (2.3) and (2.4) for each simulation. The corresponding plots of actual significance against nominal significance are shown in Figure 2.

Figure 2 indicates that both the mean adjustment and the PSS adjustment, using the *true* values of  $H$  and  $J$ , work well in this case, giving adjusted statistics which have a distribution close to  $\chi_2^2$ . Surprisingly the PSS adjustment does not perform quite as well as the mean adjustment.

Plots similar to Figure 2 for different values of  $\gamma_0$  (given in the Appendix) suggest that the adjustments work well for all  $\gamma_0 \geq 0.1$ . It is unclear whether the poor behaviour for  $\gamma = 0.05$  represents a genuine failure of the adjustments (due perhaps to the non-Gaussian distribution of the composite score) or is an artefact of the coarse quadrature algorithm for maximising the pseudolikelihood.

Figure 1 shows that when  $H$  and  $J$  must be estimated from observations, the  $\chi^2$  test based on the adjusted statistics is slightly conservative over the usual range of significance levels. At a nominal significance of 0.05 the true probability of Type I



**Figure 2:** Counterpart of Figure 1 using ‘true’ values for the matrices  $H$  and  $J$ .

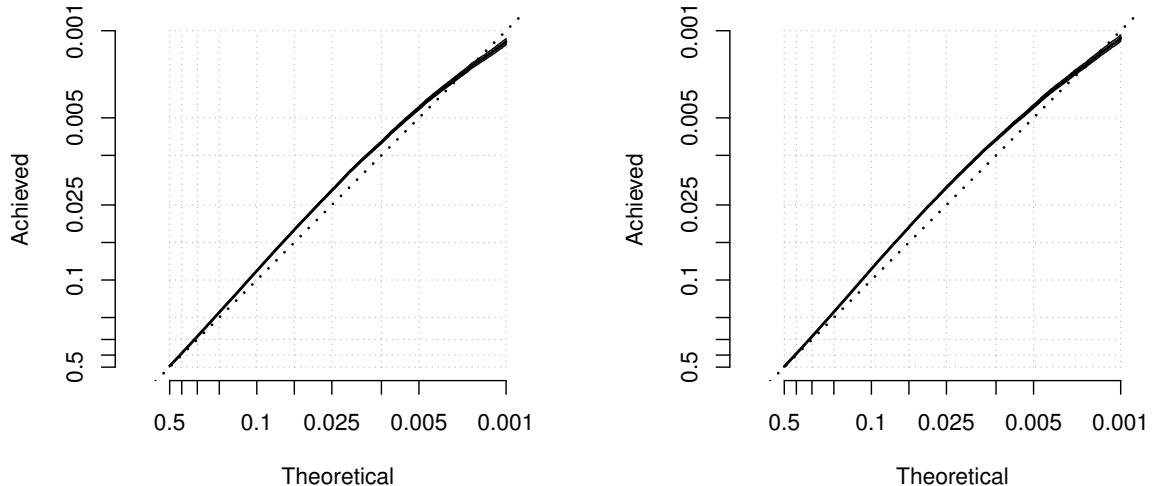
error is about 0.035. Similar plots for different values of  $\gamma_0$  suggest that there is a “breakdown” value  $\alpha^\dagger$  such that the test with nominal significance level  $\alpha$  is approximately exact or only slightly conservative when  $\alpha \geq \alpha^\dagger$ , where  $\alpha^\dagger$  increases as  $\gamma_0$  decreases: for  $\gamma_0 = 0.4, 0.3, 0.2, 0.1$  the breakdown values are about  $\alpha^\dagger = 0.005, 0.01, 0.025, 0.05$  respectively.

## 5.2 Composite null hypothesis

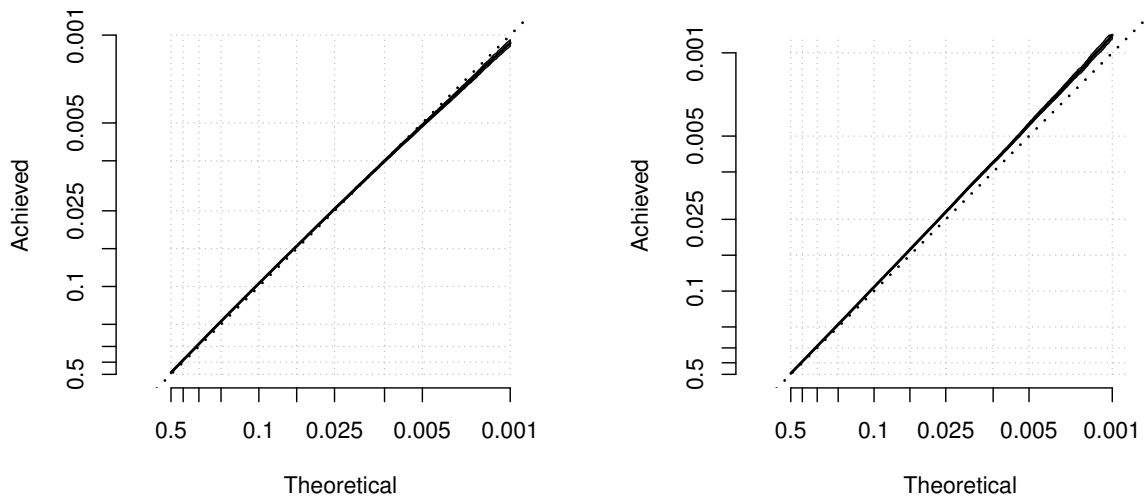
In each of the experiments described above, we also tested the composite null hypothesis  $\mathcal{H}_0 : \gamma = \gamma_0$ , in which the first order parameter  $\beta$  is treated as a nuisance parameter. For each simulated point pattern  $\mathbf{x}$  the null hypothesis was fitted by maximum pseudolikelihood (optimising  $\beta$  for fixed  $\gamma = \gamma_0$  and  $r = 0.05$ ) yielding  $\check{\theta} = (\check{\beta}, \gamma_0)$ . We then computed the pseudolikelihood score  $U(\check{\theta}, \mathbf{x})$  and the estimates  $H(\check{\theta}, \mathbf{x})$  and  $J(\check{\theta})$  using (3.5) and (3.10)–(3.11). The adjusted pseudolikelihood ratio test statistic  $\Lambda^*$  was then computed as described in section 4.2. Referring this statistic to the  $\chi_1^2$  distribution gave the adjusted  $p$ -value  $p^* = \mathbb{P}\{\chi_1^2 > \Lambda^*\}$ .

Figure 3, analogous to Figure 1, compares the empirical cumulative distribution function of  $p^*$  (based on 1 million values) against the uniform distribution, on a logarithmic scale, for the case  $\gamma_0 = 0.5$ . Performance is comparable to the case of a simple null hypothesis. Similar figures for different values of  $\gamma_0$  (given in the Appendix) show the same “breakdown” pattern as described above for the simple null hypothesis.

Figure 4, analogous to Figure 2, shows the behaviour of the adjusted pseudolikelihood ratio test statistic using the *exact* values of composite information and sensitivity estimated from 1 million simulations. Again the conclusion is that these adjustments work correctly when the moments are known exactly.



**Figure 3:** Logarithmic plot of actual significance against nominal significance for the  $\chi^2$  test of the composite null hypothesis  $H_0 : \gamma = 0.5$  based on the adjusted composite likelihood ratio test statistic, with matrices  $H$  and  $J$  estimated from data. *Left:* first moment matching adjustment. *Right:* Pace-Salvan-Sartori adjustment. Thickness of curve indicates 95% confidence interval for true significance.



**Figure 4:** Counterpart of Figure 3 using ‘true’ values for the matrices  $H$  and  $J$ .

## 6 Discussion

Our experiments used the classical point process pseudolikelihood of Besag [7]. The maximum pseudolikelihood estimator of  $\theta$  is often substantially biased [5, 33] and yet the proposed adjustments perform tolerably well.

Composite likelihoods which give better estimators of  $\theta$ , such as the logistic likelihood (Section 3.2), should also give better performance in the adjusted composite likelihood ratio test. This remains to be investigated.

The method developed here for Gibbs point process models should also extend to Cox and Neyman-Scott point processes using the Palm likelihood [8, 9, 10] or second order composite likelihood [11]. To our knowledge, variance estimators are not yet available for these composite likelihoods.

Many point process models have constraints on the parameter vector  $\theta$ . A famous example is the Strauss process [28, 29] which is well-defined only when  $\gamma \leq 1$ . Our analysis assumes no constraints on  $\theta$ ; further work is needed for constrained maximum composite likelihood.

## Acknowledgements

We thank Professor Luigi Pace for helpful advice. This research was Supported by the Danish Council for Independent Research | Natural Sciences, grant 12-124675, “Mathematical and Statistical Analysis of Spatial Data”, and by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation.

## References

- [1] A.E. Gelfand, P.J. Diggle, M. Fuentes, and P. Guttorp, editors. *Handbook of Spatial Statistics*. CRC Press, Boca Raton, FL., 2010.
- [2] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley and Sons, Chichester, 2008.
- [3] J. Møller and R.P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton, 2004.
- [4] Y.A. Kutoyants. *Statistical Inference for Spatial Poisson Processes*. Number 134 in Lecture Notes in Statistics. Springer, New York, 1998.
- [5] M.N.M. van Lieshout. *Markov Point Processes and their Applications*. Imperial College Press, London, 2000.
- [6] B.G. Lindsay. Composite likelihood. In N.U. Prabhu, editor, *Statistical Inference from Stochastic Processes*, number 80 in Contemporary Mathematics, pages 221–239. American Mathematical Society, Providence, Rhode Island, 1988.
- [7] J. Besag. Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute*, 47:77–91, 1977.
- [8] Y. Ogata and K. Katsura. Maximum likelihood estimates of the fractal dimension for random spatial patterns. *Biometrika*, 78:463–474, 1991.
- [9] U. Tanaka, Y. Ogata, and D. Stoyan. Parameter estimation and model selection for Neyman-Scott point processes. *Biometrical Journal*, 50:43–57, 2008.
- [10] M. Prokešová and E.B. Vedel Jensen. Asymptotic Palm likelihood theory for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 65:387–412, 2013.
- [11] Y. Guan. A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101:1502–1512, 2006.

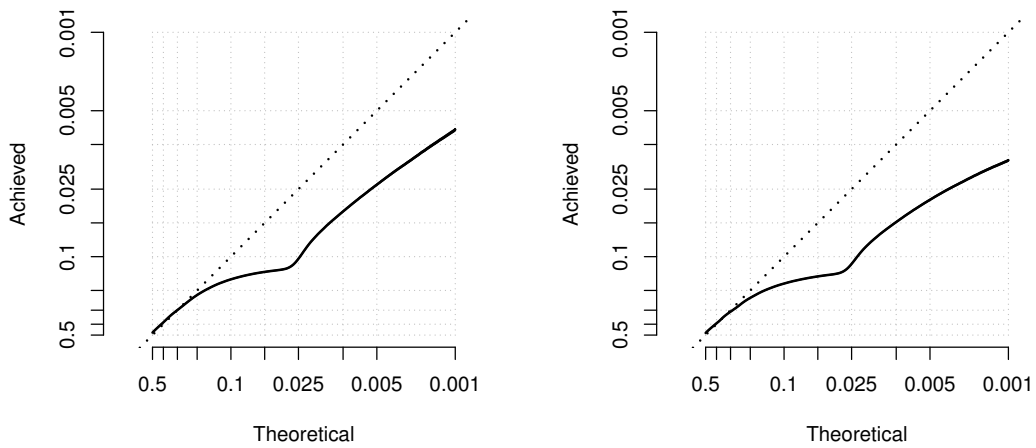
- [12] A. Baddeley and R. Turner. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42, 2005. URL: [www.jstatsoft.org](http://www.jstatsoft.org), ISSN: 1548-7660.
- [13] J.L. Jensen and J. Møller. Pseudolikelihood for exponential family models of spatial point processes. *Annals of Applied Probability*, 1:445–461, 1991.
- [14] J.L. Jensen and H.R. Künsch. On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Annals of the Institute of Statistical Mathematics*, 46:475–486, 1994.
- [15] L. Pace, A. Salvan, and N. Sartori. Adjusting composite likelihood ratio statistics. *Statistica Sinica*, 21:129–148, 2011.
- [16] A. Baddeley, J. Møller, and A.G. Pakes. Properties of residuals for spatial point processes. *Annals of the Institute of Statistical Mathematics*, 60:627–649, 2008.
- [17] J.F. Coeurjolly and E. Rubak. Fast covariance estimation for innovations computed from a spatial Gibbs point process. *Scandinavian Journal of Statistics*, 40:669–684, 2013.
- [18] A. Baddeley, J.-F. Coeurjolly, E. Rubak, and R. Waagepetersen. Logistic regression for spatial Gibbs point processes. *Biometrika*, 101(2):377–392, 2014.
- [19] J.T. Kent. Robust properties of likelihood ratio tests. *Biometrika*, 69:19–27, 1982.
- [20] A. Rotnitzky and N.P. Jewell. Hypothesis testing of regression parameters in semi-parametric generalized linear models for cluster correlated data. *Biometrika*, 77:485–497, 1990.
- [21] G. Molenberghs and G. Verbeke. *Models for Discrete Longitudinal Data*. Springer, New York, 2005.
- [22] A. Baddeley and R. Turner. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics*, 42(3): 283–322, 2000.
- [23] H.-O. Georgii. *Canonical Gibbs Measures*. Number 760 in Lecture Notes in Mathematics. Springer, Berlin, 1979.
- [24] X.X. Nguyen and H. Zessin. Integral and differential characterizations of the Gibbs process. *Mathematische Nachrichten*, 88:105–115, 1979.
- [25] J. Mecke. Stationäre zufällige maße auf lokalkompakten abelschen gruppen. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 9:36–58, 1967.
- [26] D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications*. John Wiley and Sons, Chichester, second edition, 1995.
- [27] M. Clyde and D. Strauss. Logistic regression for spatial pair-potential models. In A. Possolo, editor, *Spatial Statistics and Imaging*, volume 20 of *Lecture Notes - Monograph series*, chapter II, pages 14–30. Institute of Mathematical Statistics, Hayward, Calif., 1991. ISBN 0-940600-27-7.

- [28] D.J. Strauss. A model for clustering. *Biometrika*, 63:467–475, 1975.
- [29] F.P. Kelly and B.D. Ripley. A note on Strauss’s model for clustering. *Biometrika*, 63:357–360, 1976.
- [30] A. Baddeley and G. Nair. Fast approximation of the intensity of Gibbs point processes. *Electronic Journal of Statistics*, 6:1155–1169, 2012.
- [31] A. Baddeley and G. Nair. Approximating the moments of a spatial point process. *Stat*, 1(1):18–30, 2012. doi: 10.1002/sta4.5.
- [32] K.K. Berthelsen and J. Møller. A primer on perfect simulation for spatial point processes. *Bulletin of the Brazilian Mathematical Society*, 33:351–367, 2002.
- [33] P.J. Diggle, T. Fiksel, P. Grabarnik, Y. Ogata, D. Stoyan, and M. Tanemura. On parameter estimation for pairwise interaction processes. *International Statistical Review*, 62:99–117, 1994.

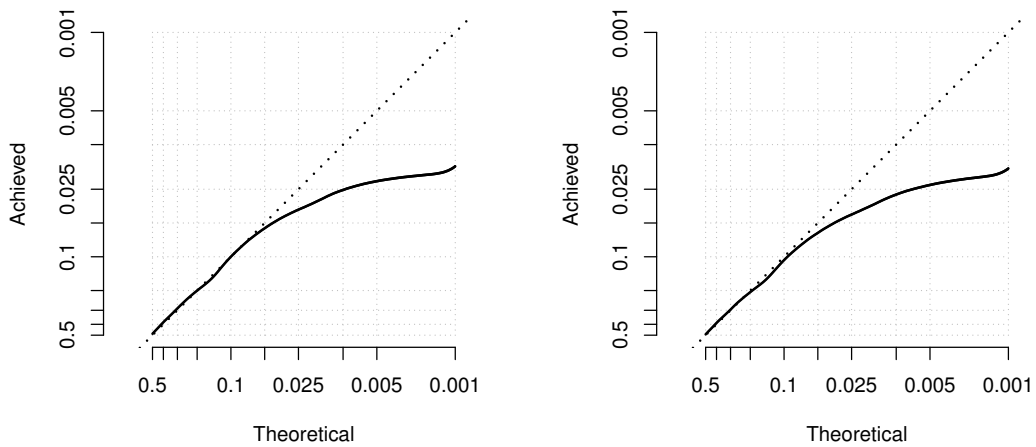
# Appendix

## A Significance plots

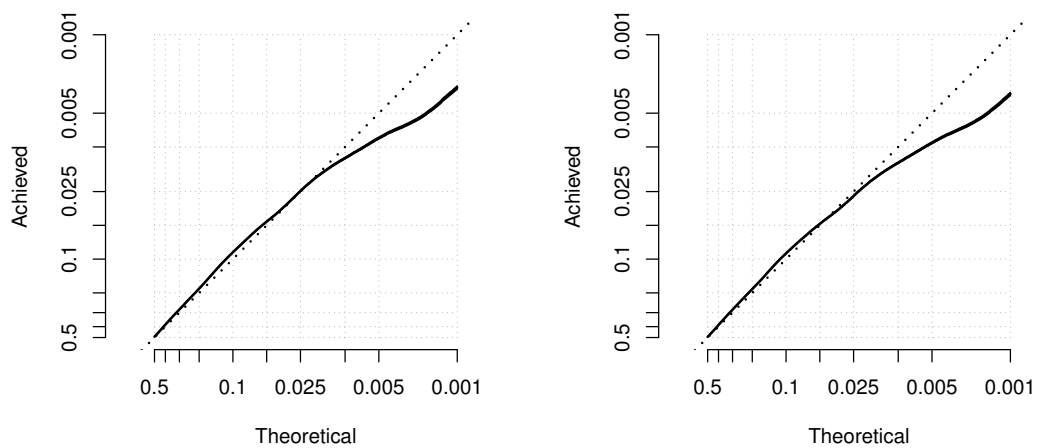
Below follows a complete list of plots of actual significance against nominal significance for the  $\chi^2$  test. The thickness of the solid curve indicates the 95% confidence interval for the true significance level (based on the asymptotic normal distribution of the proportions in this large experiment). All plots here are for the approximate test (with  $H$  and  $J$  estimated from data).



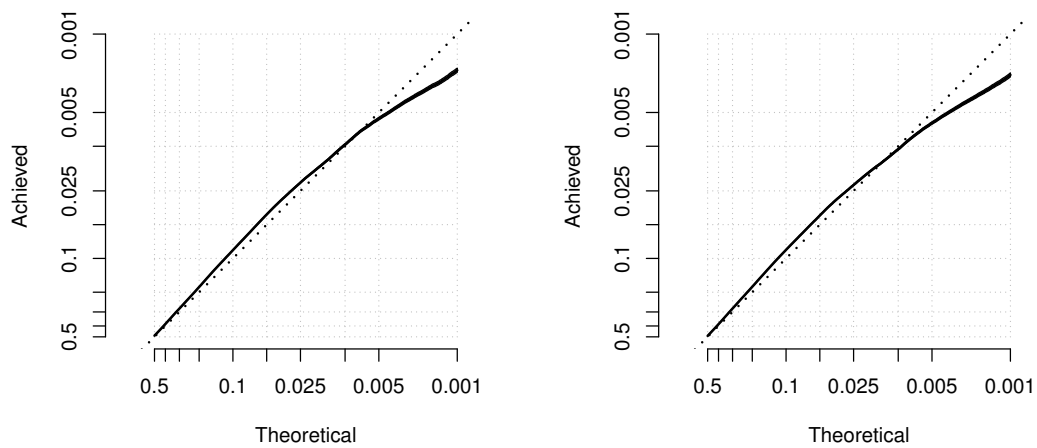
**Figure 5:** Simple,  $\gamma_0 = 0.05$



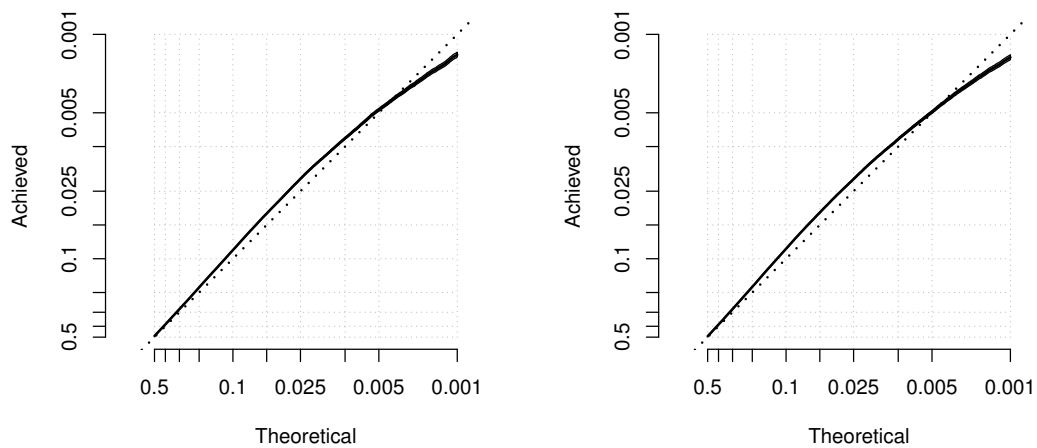
**Figure 6:** Simple,  $\gamma_0 = 0.10$



**Figure 7:** Simple,  $\gamma_0 = 0.20$

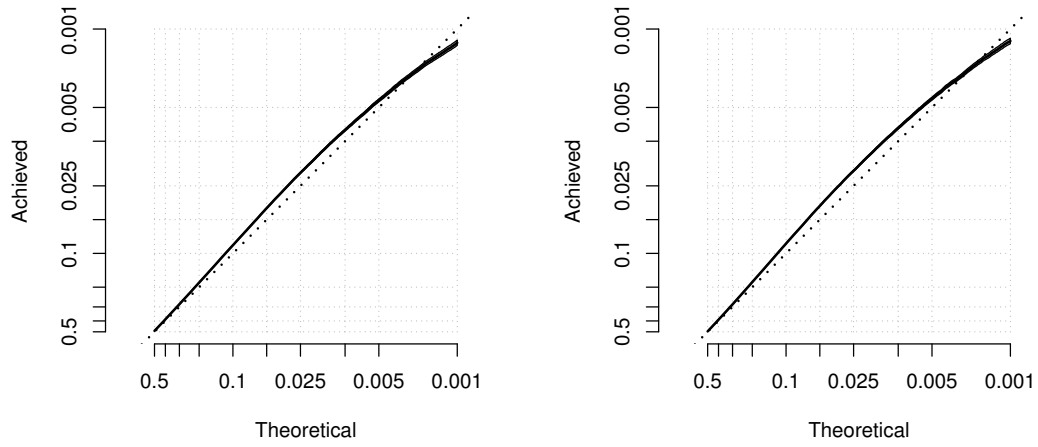


**Figure 8:** Simple,  $\gamma_0 = 0.30$

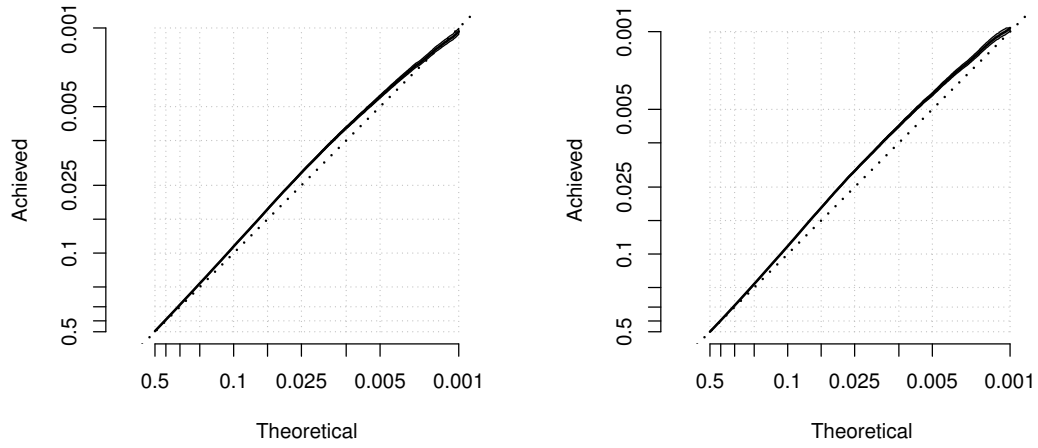


**Figure 9:** Simple,  $\gamma_0 = 0.40$

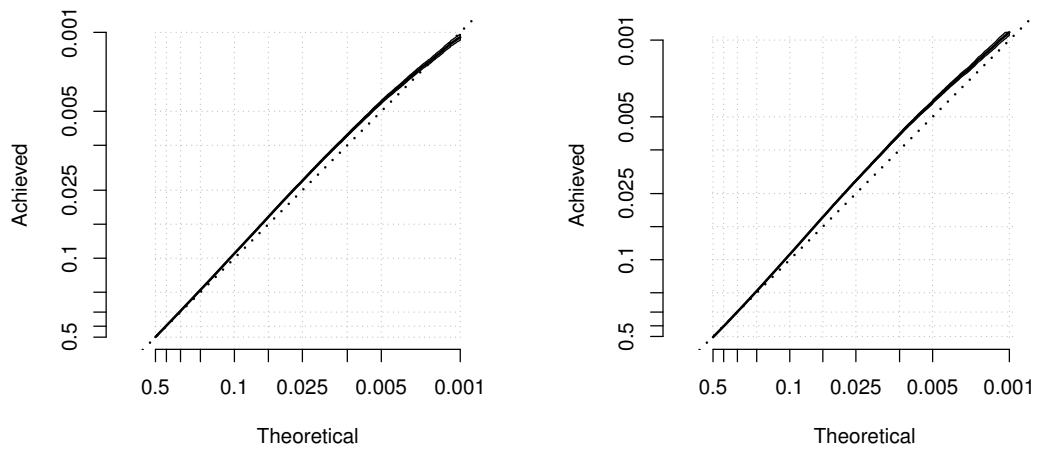




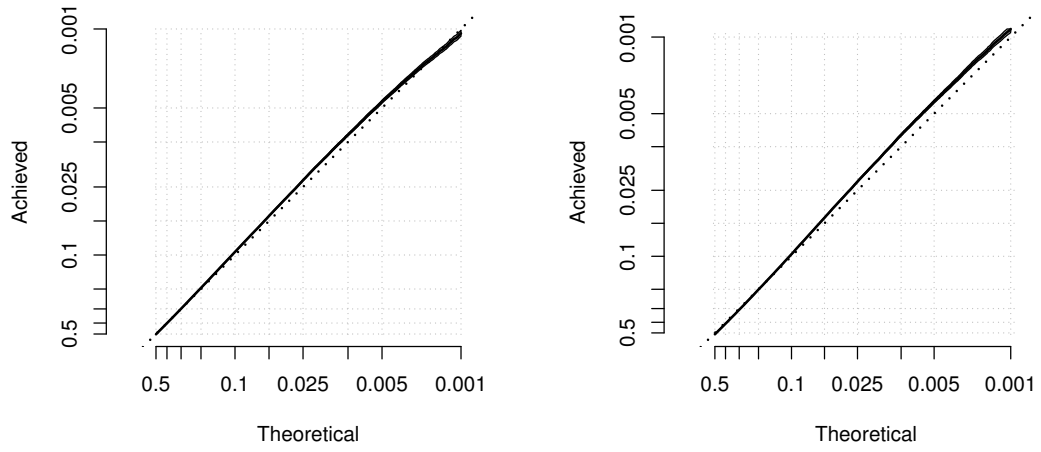
**Figure 10:** Simple,  $\gamma_0 = 0.50$



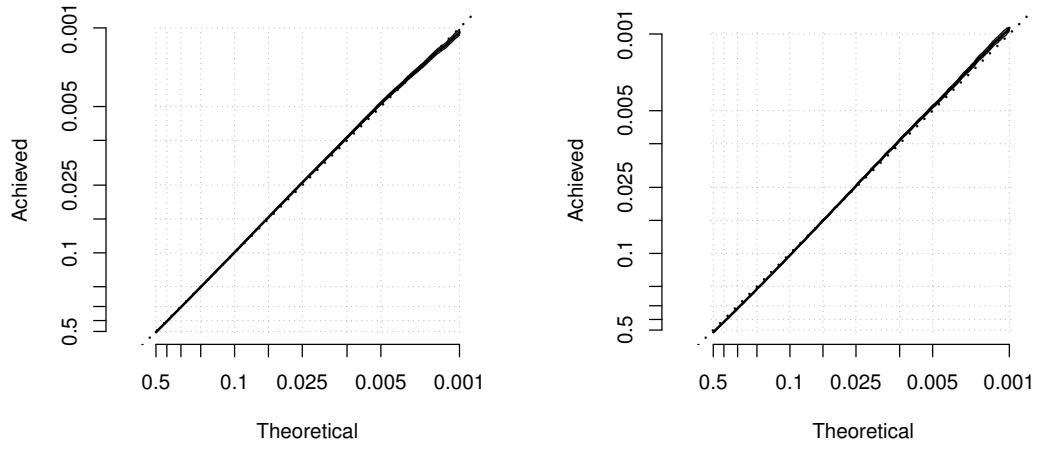
**Figure 11:** Simple,  $\gamma_0 = 0.60$



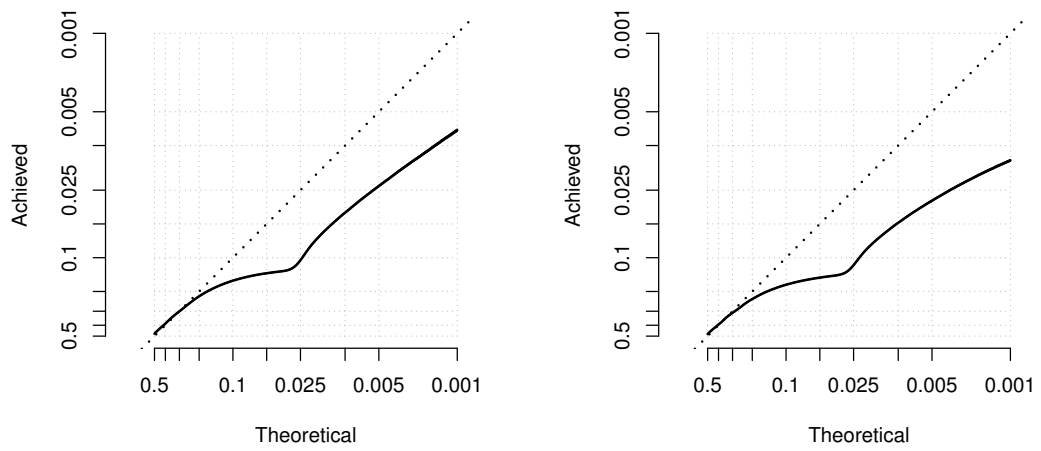
**Figure 12:** Simple,  $\gamma_0 = 0.70$



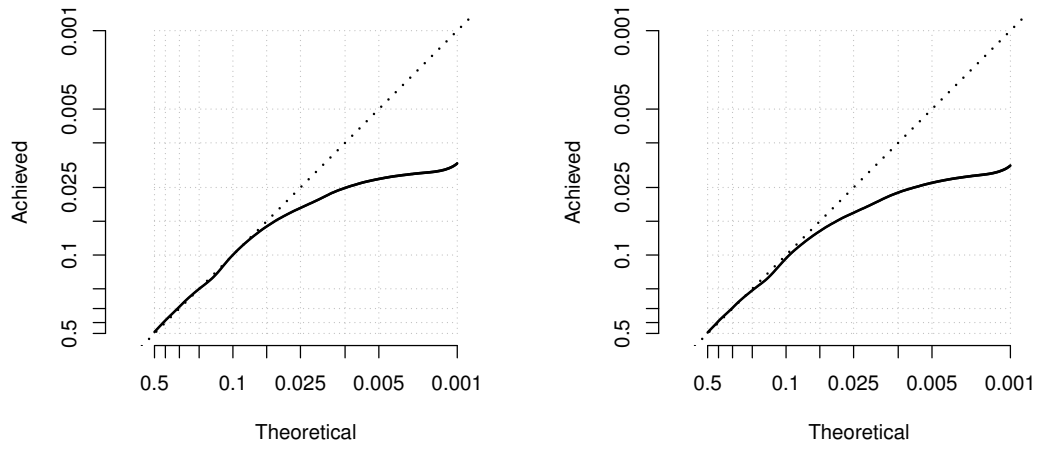
**Figure 13:** Simple,  $\gamma_0 = 0.80$



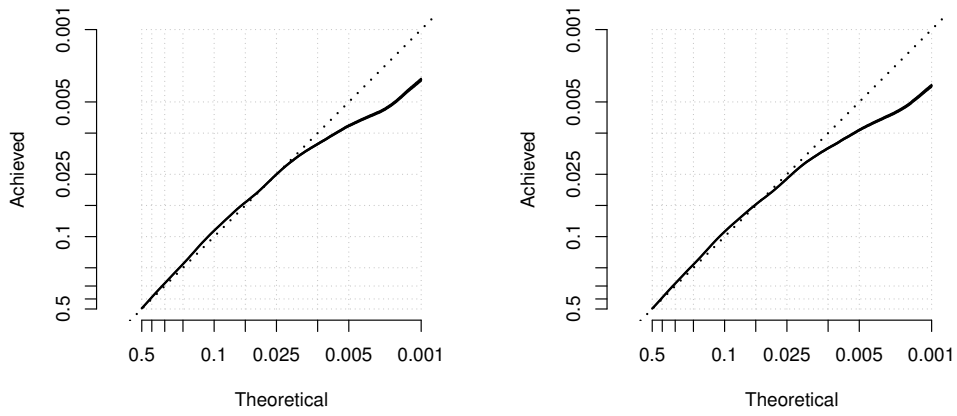
**Figure 14:** Simple,  $\gamma_0 = 0.90$



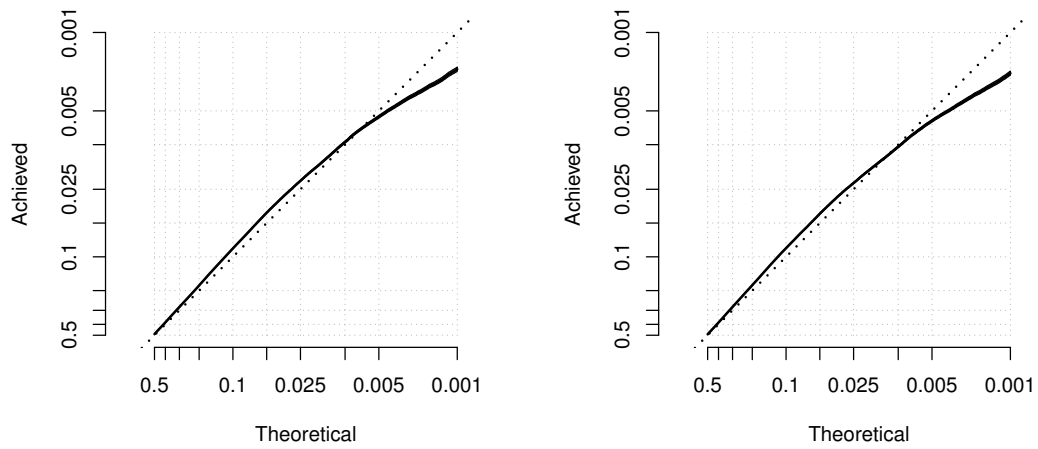
**Figure 15:** Composite,  $\gamma_0 = 0.05$



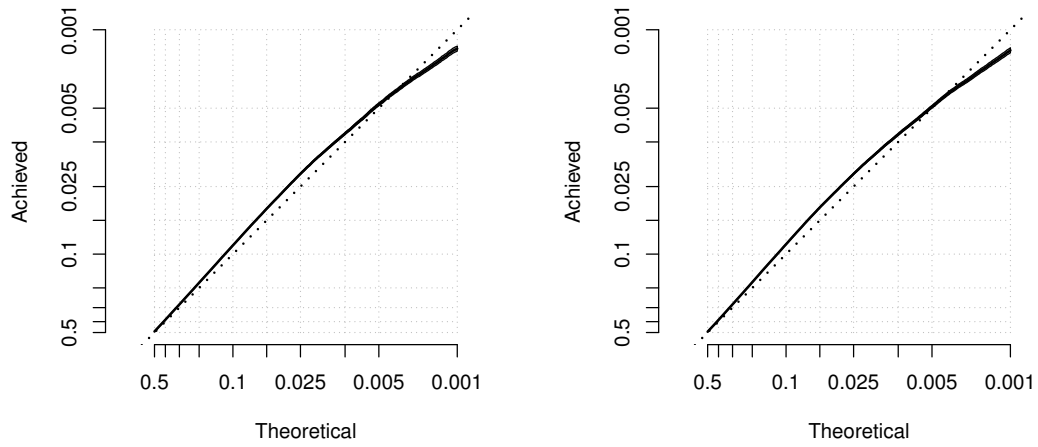
**Figure 16:** Composite,  $\gamma_0 = 0.10$



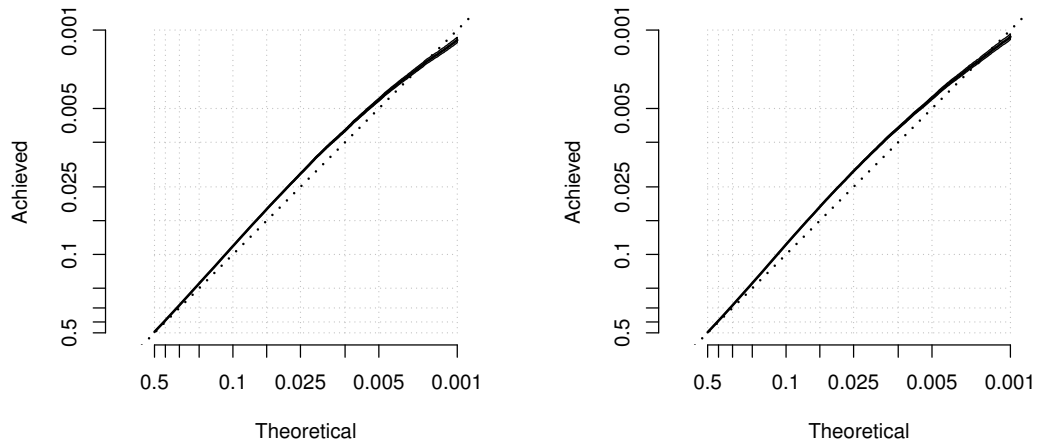
**Figure 17:** Composite,  $\gamma_0 = 0.20$



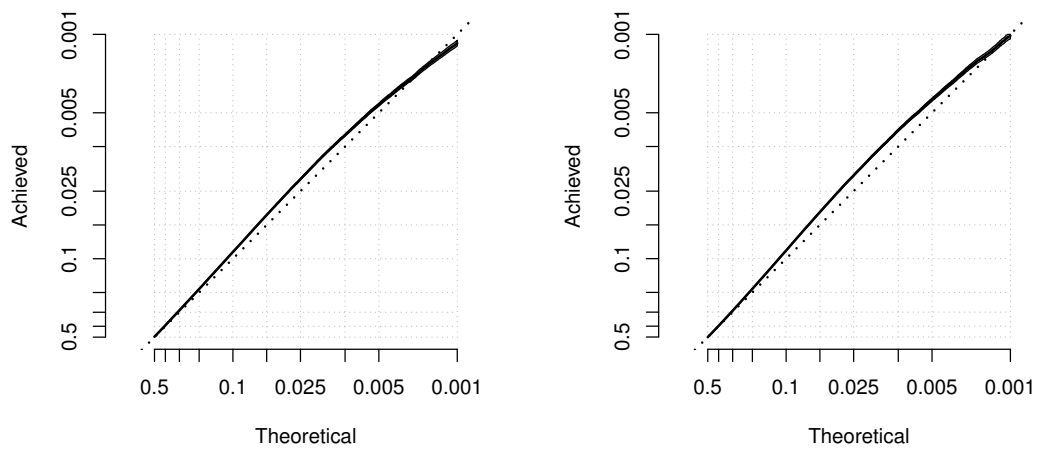
**Figure 18:** Composite,  $\gamma_0 = 0.30$



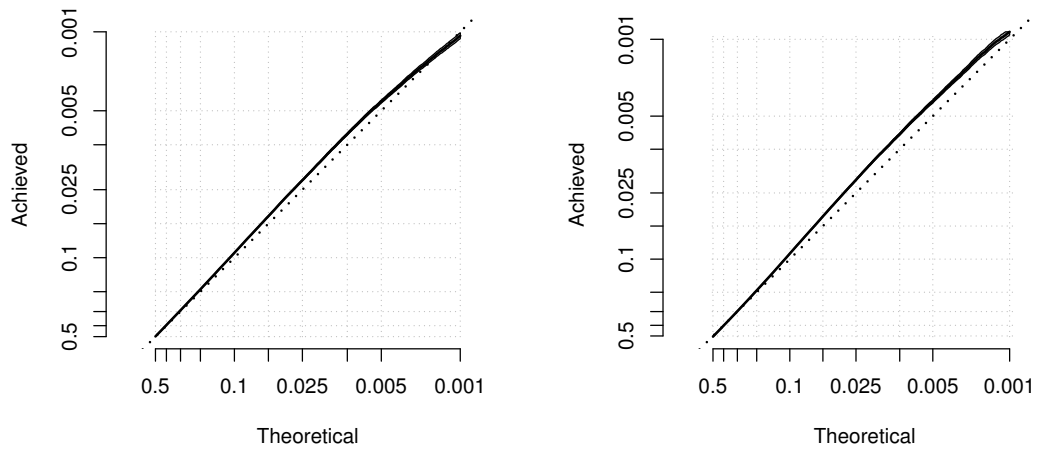
**Figure 19:** Composite,  $\gamma_0 = 0.40$



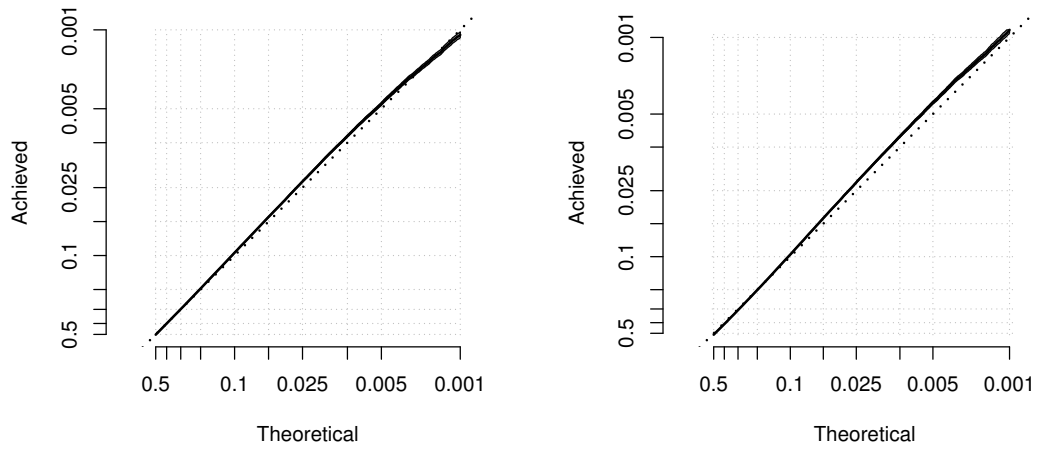
**Figure 20:** Composite,  $\gamma_0 = 0.50$



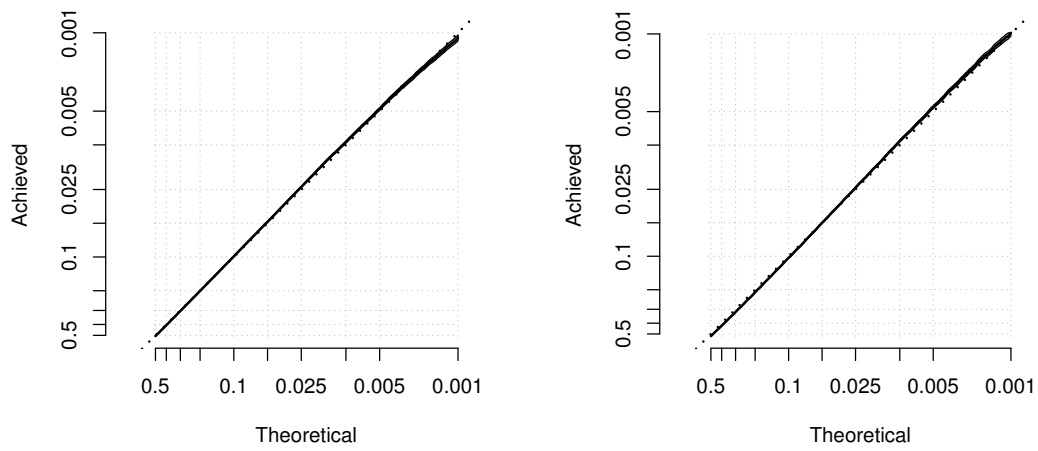
**Figure 21:** Composite,  $\gamma_0 = 0.60$



**Figure 22:** Composite,  $\gamma_0 = 0.70$



**Figure 23:** Composite,  $\gamma_0 = 0.80$



**Figure 24:** Composite,  $\gamma_0 = 0.90$