A Nonparametric Analogue to POSTHOC Estimates for Exploratory Data Analysis

Robert H. Leary and Jason Chittenden

Pharsight® Corporation, Cary, North Carolina 27518, USA

Introduction

The use of FO or FOCE POSTHOC eta values is a commonly used approach for exploratory data analysis of possible covariate relationships (e.g., regressing post hoc estimates of a clearance or volume of distribution against weight or against other POSTHOCs) and distributional features. However, as shown in [R. Savic and M. Karlsson, PAGE 2007, abst. 1087], shrinkage effects in the sparse data case may hide or distort an actual dependence or correlation, possibly rendering such analyses ineffective or even misleading. Recently NONMEM® VI has added a relatively simple nonparametric capability NONP in which the nonparametric maximum likelihood (NPML) distribution is approximated by a discrete distribution with support points fixed at the POSTHOC estimates from a preliminary parametric FO or FOCE analysis. NPML optimization is performed only over the associated probabilities on the support points. Due to shrinkage and/or excessive POSTHOC correlations, the fixed POSTHOC supports may be badly placed relative to the supports in a NPML distribution that has also been optimized with respect to support point positions. Here we consider the use of the mean of the individual nonparametric posterior distributions as a nonparametric analogue to parametric POSTHOCs in exploratory data analysis, using both NONP as well as full NPML optimization.

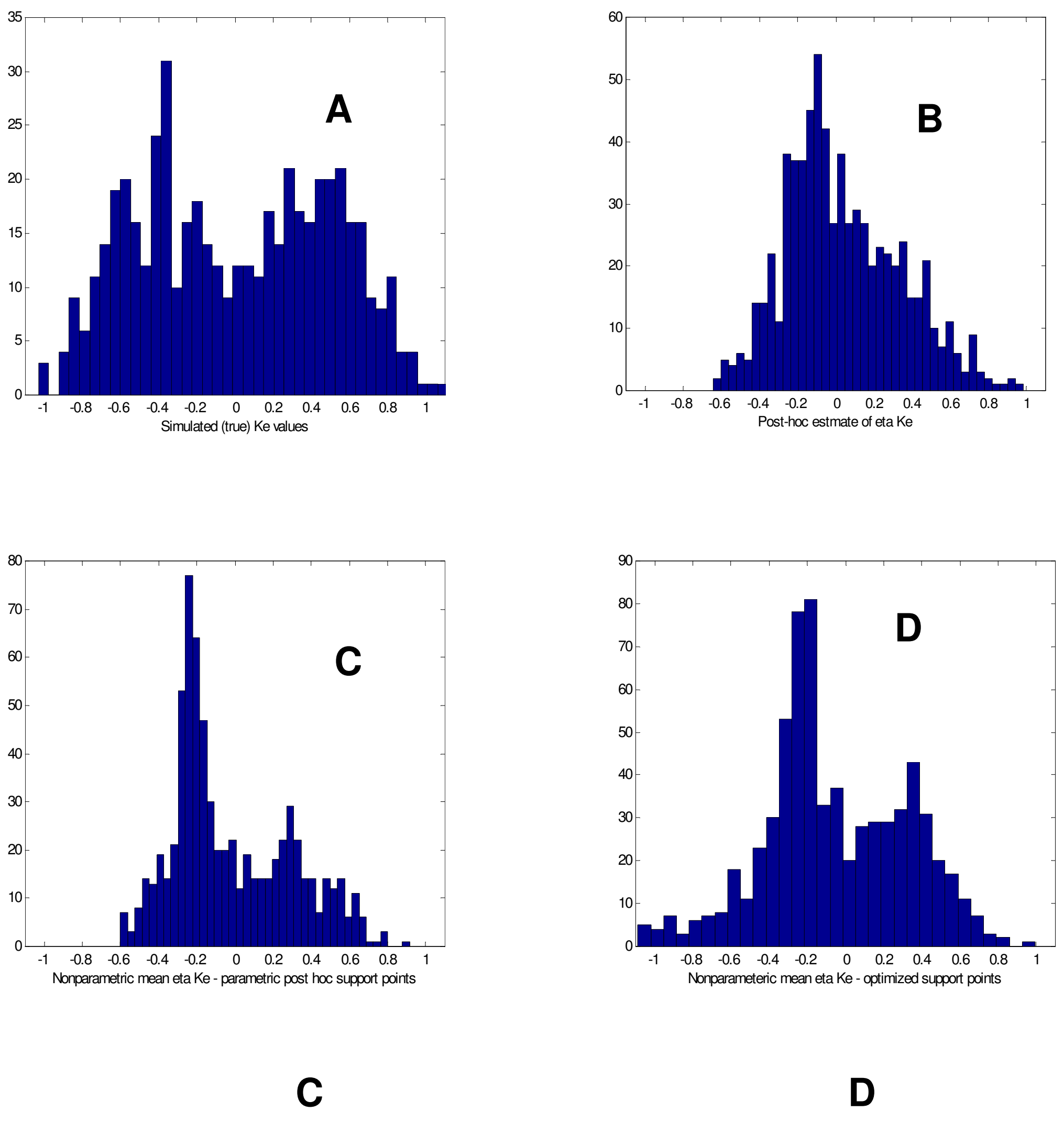
Parametric POSTHOCS can be poor nonparametric candidate supports

Fig 1A shows the true observed bimodal distribution for 700 subjects (ETA_Ke ~ 1/2 N(-0.41, 0.0625) + ½N(0.41, 00625), eta_V~N(0, 0.0625), Ke = exp(eta_Ke), V=exp(eta_V), from a simulated linear one-compartment IV bolus NLME model

DV = dose*exp(-Ke*time)/V *exp(eps) , eps= 0.1

while Fig 1B shows the distribution of the post hoc estimates of eta_Ke after an FOCE fit using a normality assumption for the random effects . Note that the bimodality in eta_Ke has been completely masked by the shrinkage phenomenon. Fig 1C shows the bimodal but still too narrow distribution of the means of the individual subject posterior distributions from an NP fit with parametric FOCE post hoc supports, while Fig 1D shows the wider corresponding distribution of the individual posterior means for a fully optimized NP fit . The nonparametric 2LL value for D is higher than for C by 24.506.

Fig. 1



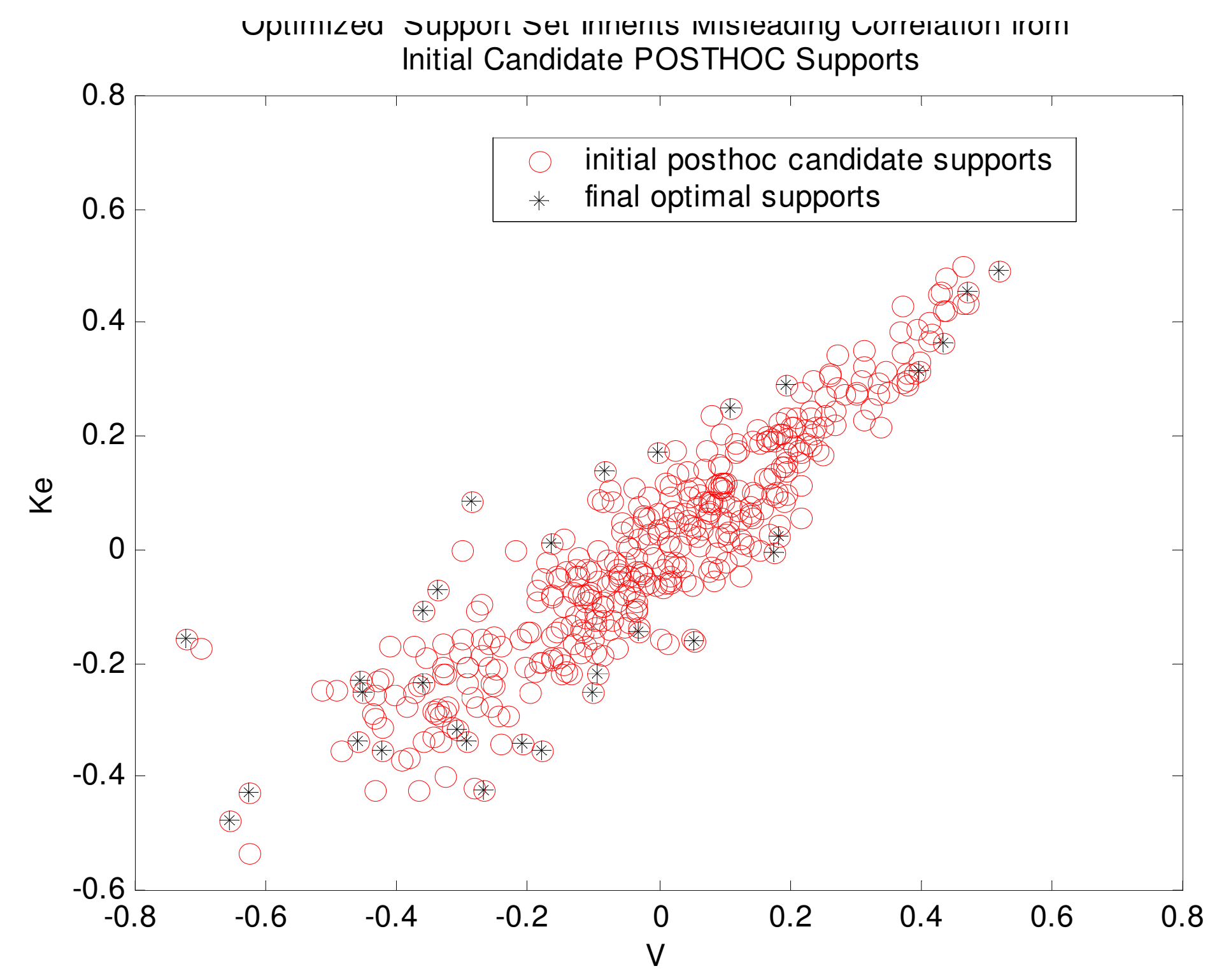


Fig. 2. An FOCE fit to a simple IV bolus Ke-V model with sparse (2 sample times) data (Nsub = 400) simulated from a true multivariate normal random effects distribution with diagonal Ω shows highly correlated (ρ = 0.91) POSTHOC etaKe and etaV values, even thought the true underlying eta values are independent. An NONP analysis using the excessively correlated POSTHOCS as the candidate supports selects final supports on the periphery and reduces the apparent POSTHOC correlation using NP individual posterior distribution means) to ρ = 0.85, but overall the high correlation of the initial candidate supports cannot be overcome by the nonparametric analysis. In contrast, in the full nonparametric analysis with optimized supports shown above right in Fig. 3, the correlation of the NP individual posterior means is ρ = 0.11.

Computation of NP "POSTHOC" Individual Posterior Means

For a given set of M candidate support points, the NPML distribution P with N subjects is discrete on at most N support points and can be found as the solution to

Maximize NPML(P), where

$$NPML(P) = \sum_{I=1}^{N} \log\left(\sum_{J=1}^{M} L_{IJ} P_J\right)$$

$$P_J \geq 0, \sum_{J=1}^{M} P_J = 1$$

L_{IJ} are fixed conditional likelihoods determined by evaluating the residual error model at all support point S_J for subject I.

This is a convex optimization problem that can be rapidly and reliably solved, even with very large candidate support sets, with a central path following primal-dual algorithm. The solution consists of at most N (# subjects) support points with $P_J > 0$. The remaining points associated with $P_J = 0$ are discarded.

If the starting set of candidate support points is very rich, (e.g. a high resolution grid over the random effects space), a single application of the primal dual algorithm is sufficient to obtain a near global optimum over both probabilities and support point positions (see fig. 3 above right). Alternatively, a sequential approach of solving on successively refined sparser grids can be used.

In addition to the population distribution, an NP posterior distribution $P_J(I)$ on the same final supports for each subject I is obtained by using the I-th row of L to compute a Bayesian posterior with P_J as a prior. The mean of P_J (I) is analogous to the parametric POSTHOC values for subject I and can be used similarly for exploratory data analysis. It can be more resilient to distortions caused by shrinkage and other phenomena in the sparse data case.

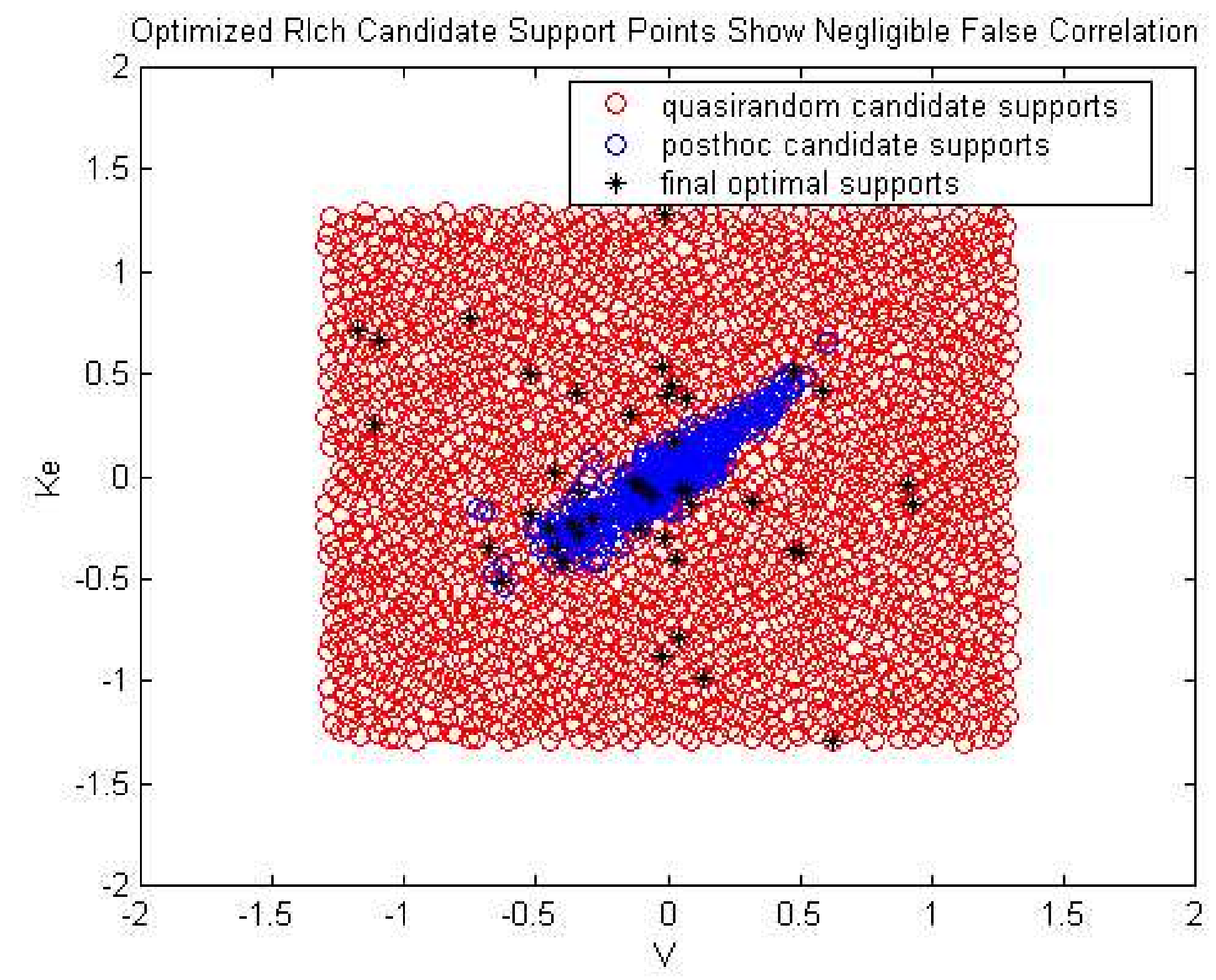


Fig. 3. In contrast to the situation in Fig. 2 (left), a fully optimized nonparametric analysis (above) using an initial dense grid of quasi-random Sobol sequence support points selects optimal supports well outside the narrow band of correlated POSTHOC values. The corresponding correlation of the NP individual distribution means is reduced to ρ = 0.11

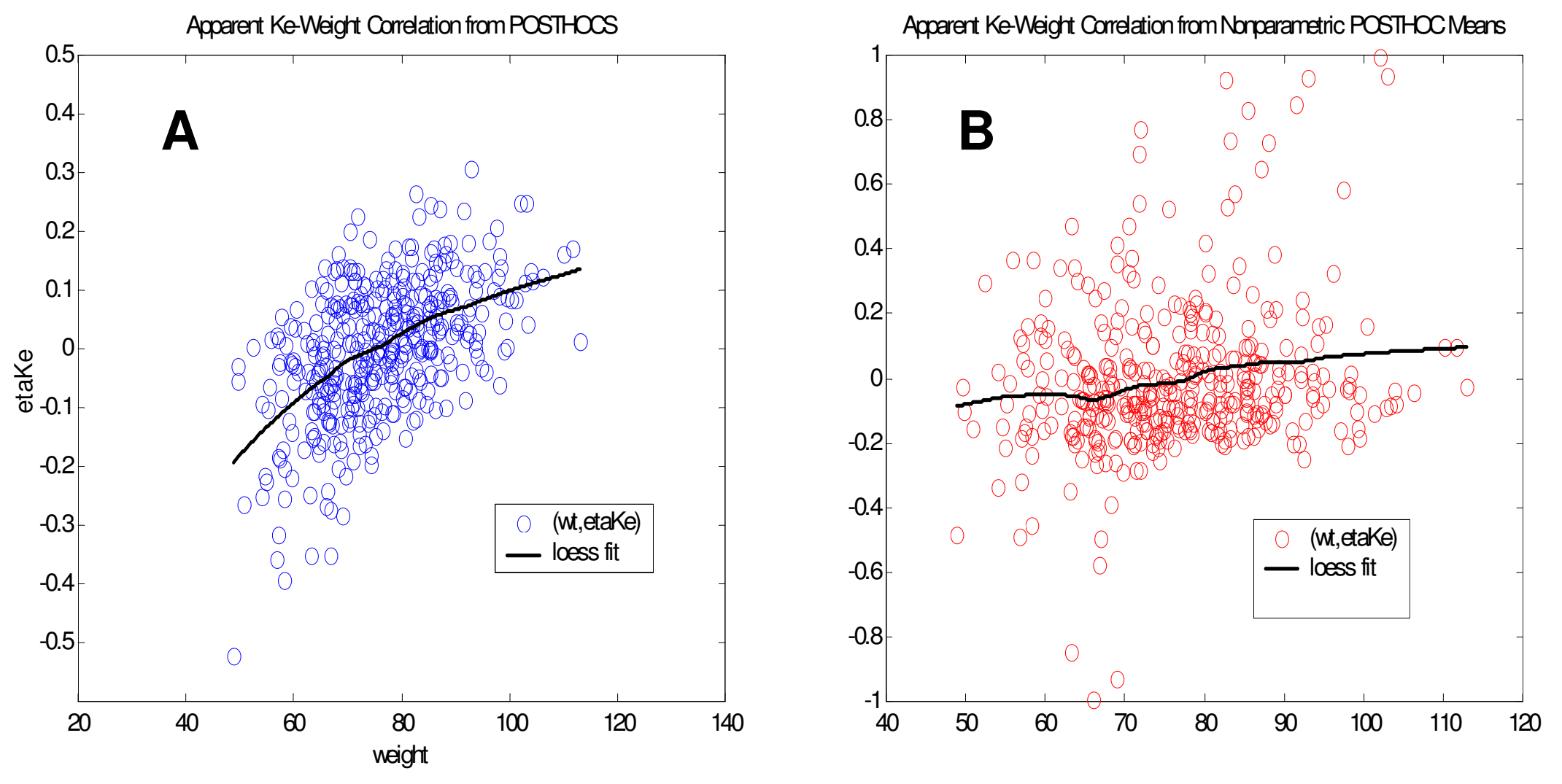


Fig.4. In a second similar simulation, a covariate relationship between V (but not Ke) and weight is introduced into the simulated data set, with etaV and etaKe remaining independent. The FOCE analysis with a base model continues to show a strong (ρ > 0.9) correlation between the POSTHOCS for etaV and etaKe. The ρ = 0.55 etaV-weight correlation is correctly identified by a POSTHOC plot (not shown) but a second spurious ρ = 0.48 etaKe-weight correlation is induced (Fig. 4A). In Fig. 4B, the spurious correlation is essentially removed (ρ =0.11) by a full NPML analysis (but not by NONP optimization over POSTHOCS).

Conclusions (sparse data case)

1) Strong (ρ > 0.9) correlations may be observed between POSTHOC values when the true underlying random effects are uncorrelated, making the POSTHOCs ineffective for exploratory data analysis

2) When one of two apparently correlated etas is also correlated with a covariate, a false eta-covariate correlation may be induced in the other

3) Means of NPML estimated individual eta distributions may reflect true eta-eta and eta-covariate correlations more accurately than parametric POSTHOCS , but

4) A full NPML estimation procedure requires optimization over both probabilities and support point positions – simple NONMEM® NONP optimization over POSTHOC supports will inherit the poor exploratory characteristics of the parametric POSTHOCs.

5) NPML posterior individual means may also reveal features such as bimodality in the population distribution that are hidden in parametric analyses.

Trademarks

NONMEM, Globomax, Inc.

Pharsight, Pharsight Corporation

Contact

Robert H. Leary, PhD
Pharsight Corporation
Ph: +1 919.852.4625
Email: bleary@pharsight.com