Published in final edited form as:

Behav Genet. 2013 January ; 43(1): 85–96. doi:10.1007/s10519-012-9573-7.

A GENERALIZED DEFRIES-FULKER REGRESSION FRAMEWORK FOR THE ANALYSIS OF TWIN DATA

Laura C. Lazzeroni and Amrita Ray

Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, 401
Quarry Road, Stanford, CA 94305-5723, USA

Laura C. Lazzeroni: lazzeroni@stanford.edu

INTRODUCTION

Genome-wide association studies (GWAS) may identify one or more genetic variants associated with a complex trait, but the collective explanatory power of those variants is often weak. The variants may account for only a small part of the disease risk or trait variance that prior heritability estimates suggest is due to genetic factors. This phenomenon is called “missing heritability” (Visscher et al. 2008). Several factors could explain why GWAS might consistently fail to detect sources of heritability that, in fact, exist (Manolio et al. 2009).

An alternative, which we call “misestimated heritability”, would be to attribute the discrepancy to the unreliability of heritability estimates usually obtained from twin studies. Methods for the analysis of twin studies, as currently implemented, have limitations that can potentially lead to biased estimates and computational convergence problems. Moreover, most methods do not provide reliable ways to test data for possible violations of key assumptions of the twin model. Nor do they provide reliable confidence intervals or other measures of estimator uncertainty, which studies therefore rarely report.

DeFries and Fulker (1985) introduced a computationally stable regression model for twin data that avoids the biases of some other estimation methods. However, DeFries-Fulker (DF) regression lacks an implementation that formally incorporates covariates, a potential source of confounding if ignored. In this paper, we propose a generalized DF (GDF) regression model that incorporates covariates. We use the DF framework to construct tests of key twin model assumptions and confidence intervals for parameter estimates. We perform a simulation to compare test, estimator and confidence interval performance of GDF regression to the widely used, mixed effects model. We illustrate GDF regression on data from a study of bronchopulmonary dysplasia in premature twins. In order to motivate and explain the GDF model, this paper also addresses some conceptual issues. (1) We develop a novel triangular representation of the twin parameter space and consider the implications, advantages and disadvantages of imposing non-negativity constraints on the variance components, a key difference between mixed effects and GDF models. (2) We develop a new, family-resemblance interpretation of twin models that arises naturally from the GDF model and show its relationship to the recurrence risk for binary traits. (3) We discuss possible future extensions of the GDF strategy for generalized linear models, truncated or censored data and gene-by-environment interactions on the variance components.

Conflict of Interest: The authors have no conflicts of interest to declare.

THE CLASSICAL ACDE MODEL

Twin studies were first proposed by Galton (1875) as a way to study the relative contributions of genes and environment to a disease or other complex trait. They rely on the fact that the genetic similarity of monozygotic (MZ) twins is greater than that of dizygotic (DZ) twins. It follows that MZ twins should be more similar than DZ twins with respect to genetically transmitted or heritable traits. Fisher (1918) proposed the first variance component models, partitioning the variance of a trait according to the classical ACDE twin model still in use today.

Twin study data provide direct estimates of two parameters related to family aggregation or heritability. For continuous data, these estimates are the two separate within-pair correlations for MZ and DZ twin pairs. Others measures of similarity, such as concordance rates, are also possible. Heritability estimates, however, are defined in terms of the variance components of the ACDE twin model. Let γ_{MZ} and γ_{DZ} be the correlations or similarity measures for MZ and DZ pairs, respectively. The relationship between the correlations and the ACDE parameters is given by $\gamma_{MZ} = A + C + D$ and $\gamma_{DZ} = 0.5A + C + cD$. A refers to correlation due to Additive genetic effects, which is defined to be the narrow-sense heritability under the ACDE model. C refers to correlation due to family environmental effects shared in Common. D refers to Dominance or other interactive genetic effects, which may involve multiple loci. E denotes independent random variation due to an individual “Error” term. MZ pairs are considered genetically identical and share all A and D effects. DZ twins are only 50% identical by descent (IBD) on average and share only half the additive genetic effects. The constant c , the proportion of the D effects shared within DZ twin pairs, is unknown and can depend on a complex combination of interactive allelic effects, linkage disequilibrium and linkage. A key assumption, not addressed here, is that the sharing of the common environmental effects within families, C , is the same for MZ and DZ pairs (Hettema et al. 1995; Kendler 1993). Other complications, not consistent with the ACDE model, may also exist. For example, non-random mating (inbreeding or population stratification) can mimic the appearance of family environmental effects by increasing the similarity of DZ twins (Martin et al. 1986; Boomsma et al., 2002; Friedman et al. 2003).

Figure 1 represents the ACDE parameter space in terms of the estimable correlations γ_{MZ} and γ_{DZ} . Two inequality constraints, $\gamma_{DZ} \geq 0$ and $\gamma_{MZ} \geq \gamma_{DZ}$, follow from the fact that A , C , D and c are all non-negative by assumption. The E or null model, in which there are no environmental or genetic family effects, is the point (0, 0) at the corner of the space. The CE model includes only family environmental effects and corresponds to the horizontal line at the lower boundary of the ACDE triangle where $\gamma_{MZ} = \gamma_{DZ}$. The AE model includes only additive genetic effects and corresponds to the interior diagonal line where $\gamma_{MZ} = 2\gamma_{DZ}$. The ACE model, under which $\gamma_{MZ} \leq 2\gamma_{DZ}$, is the region below the diagonal. The ADE model, under which $\gamma_{MZ} \geq 2\gamma_{DZ}$, is the region above the diagonal. Any pair of observed correlation estimates, γ_{MZ} and γ_{DZ} , will yield a single point in the triangle, which may be consistent with either the ACE or ADE model, but not both. Thus, C and D cannot be simultaneously statistically significant, although both types of effects may be present in reality. Moreover, it is unnecessary to fit the ACE and ADE models separately from the ACDE and AE models.

Estimates of the ACDE, CE, AE and E models can be used to construct hypothesis tests. A valid test for familial aggregation is given by a test of either the CE or the AE model against the E model. We anticipate the AE (CE) test will have greater power if genetic (environmental) family effects dominate. A valid test for heritability or a genetic effect is given by a test of the ACDE model against the CE model. The previous hypothesis tests are one-sided. A valid test for genetic complexity, defined as the presence of family

environmental, genetic dominance or interactions or both, is given by a two-sided test of the ACDE model against the AE model. Genetic complexity may make it more difficult to identify specific genetic factors. The boundary constraints affect the implementation of these tests under the mixed effects model. Although the full ACDE model can be used to carry out hypothesis tests, the variance components A (i.e. the heritability), C and D are not identifiable (Kempthorne and Osborne 1961; Haseman and Elston 1970; Karlin et al. 1983; Williams 1993). Parameter estimation requires additional assumptions. For example, $2(\gamma_{MZ} - \gamma_{DZ})$ is an estimate of A under the ACE model when D is assumed to be zero.

The triangle constraints

Mixed effects models, widely used in the analysis of twin data, naturally enforce the triangle constraints on the resulting correlation estimates. DF regression does not and its estimates of γ_{MZ} and γ_{DZ} can lie outside the triangle. An important question is whether it is better for analyses of twin data to enforce the triangle constraints automatically. On the one hand, we expect constrained estimates to be more efficient when the true model is consistent with the constraints. Automatic constraints also eliminate the problems of interpretation that arise when estimated values are inconsistent with the assumed ACDE model. On the other hand, true parameter values on or near a constraint boundary can create a number of problems. Estimation algorithms can take a long time in this setting and may ultimately fail to reach convergence. Proximity to boundary constraints can yield poor standard error estimates and skewed sampling distributions. Together, these problems limit the availability of reliable confidence intervals for mixed effects models. See Carey (2005) for a thorough examination of these issues.

Allowing estimates to lie outside the constraints enables goodness-of-fit tests of the ACDE model itself. For example, observed correlations may be less than zero or the MZ correlation may be less than the DZ correlation. If the estimation method is unconstrained, it is possible to test whether such observations reflect statistically significant violations of the ACDE model. Such violations could point to data errors, outliers, problems in sample recruitment or protocol implementation. Goodness-of-fit tests are not possible when the estimation method forces the estimates to conform to the model.

For some traits, the true model may be inconsistent with the ACDE triangle. Behavioral and psychiatric studies frequently observe MZ correlations that are much more than twice the DZ correlation (e.g. Hicks et al. 2004; Young et al. 2000). This pattern can be interpreted as reflecting complex interactions among multiple genetic loci (Risch 1990). It can also be explained by additive genetic factors, offset by a negative correlations ($C < 0$) due to a shared, but dynamic family environment (Figure 2). For example, both MZ and DZ twins may compete for some limited resource, mediating variable or risk factor (food, parental attention, prenatal nutrition, monetary resources). Twins might also tend to adopt complementary or contrasting roles within the family (leader-follower, introvert-extrovert).

MODEL ESTIMATION

DF regression

DF regression is one of several methods that can be used to estimate twin pair correlations and the components of the ACDE model. The original DF model (DeFries and Fulker, 1985) regresses y_i , the quantitative phenotype of twin i , on y_j , the phenotype of co-twin j , without covariates. Using different notation in the formulation of Rodgers and Kohler (2005), the DF regression equation for a single twin is

$$y_i = b + f_1(y_j - b) + f_2 R(y_j - b) + e_i, \quad (1)$$

where R is the expected IBD proportion ($R = 1$ for MZ pairs and $1/2$ for DZ pairs); b is the sample mean of y and e_i is a random error term with variance σ_e^2 . Error terms are assumed independent. When data are unselected (i.e., there are no probands), each twin pair is entered in the data set twice, once with y_i and again with y_j as the dependent variable, a practice known as “dual” or “double” entry. Standard linear regression is applied to the dual entry data set to obtain estimates of f_1 and f_2 . DF estimates of f_1 and f_2 are unbiased estimates of C and A, respectively, in the ACE model. Waller (1994) was the first to develop DF methods for models including non-additive genetic dominance. Standard errors for these estimates have been derived that take into account dual entry and the relatedness of twins from the same pair. An advantage of DF regression is its computational simplicity and speed as it relies on closed-form expressions for the estimates and standard errors. Later work has added to the theoretical basis of DF regression and extended it to other models, including logit and probit models for binary traits and tobit models for censored data (Cherny et al. 1992; Rodgers and McGue 1994; Sham et al. 1994; Kohler and Rodgers 1999; Kohler and Rodgers 2001; Purcell and Sham 2003).

A generalized DF regression model

We extend the DF regression model to include covariates, by letting x_{ik} be the value of the k^{th} of p total covariates for the i^{th} twin, with $x_{i0} = 1$. We replace the sample mean in equation (1) by the fixed covariate effects for both twins, yielding

$$y_i = \sum_{k=0}^p b_k x_{ik} + f_1(y_j - \sum_{k=0}^p b_k x_{jk}) + f_2 R(y_j - \sum_{k=0}^p b_k x_{jk}) + e_i \quad (2a)$$

$$= \sum_{k=0}^p b_k x_{ik} + (f_1 + f_2 R) \times (y_j - \sum_{k=0}^p b_k x_{jk}) + e_i, \quad (2b)$$

where the b_k s are unknown coefficients. For the generalized DF (GDF) model, we regress the phenotype of each twin on his or her own covariate values and on the co-twin's phenotype residual, given the co-twin's covariate values. Equation (2) includes all available information for estimating f_1 and f_2 . However, the marginal regression equation,

$$y_i = \sum_{k=0}^p b_k x_{ik} + e_i', \quad (3)$$

adds information for the estimation of the covariate effects. The GDF model differs from previous DF models in that the equation includes the covariate values of both twins and additional, marginal information is used to estimate the b_k s. Below, we describe a method for estimating the GDF model using dual entry data and equations (2) and (3).

The relationship between the variance components of the twin model and the coefficients, f_1 and f_2 , of the GDF (or DF) model follows from known properties of the Normal

distribution. If y_i and y_j are jointly Normally distributed with expectations $\sum_{k=0}^p b_k x_{ik}$ and $\sum_{k=0}^p b_k x_{jk}$, the conditional expectation of y_i given y_j is $\sum_{k=0}^p b_k x_{ik} + \rho(y_j - \sum_{k=0}^p b_k x_{jk})$, where ρ is the correlation of y_i and y_j (Castillo and Galambos 1989). It follows from equation (2b) that the total within-pair correlation is $\gamma_R = f_1 + f_2 R$. Thus, $\gamma_{MZ} = f_1 + f_2$ and $\gamma_{DZ} = f_1 + 0.5 f_2$.

Conversely, $f_2 = 2(\gamma_{MZ} - \gamma_{DZ}) = A + (1-c)D$ and $f_1 = 2\gamma_{DZ} - \gamma_{MZ} = C + (2c-1)D$ in terms of the ACDE twin model. In the model in Figure 2 with $D = 0$, f_1 equals $\gamma_{R=0}$ and f_2 equals $\gamma_{R=1} - \gamma_{R=0}$. If D is assumed to be zero, $f_1 = C$ and $f_2 = A$. To fit the CE or AE model, we omit f_2 or f_1 , respectively. Another consequence of the Normality assumption is that the conditional variance of y_i is $\sigma_e^2 = (1 - \gamma_R^2)\sigma_y^2$, where σ_y^2 is the unconditional variance of y .

Hypothesis Tests

Under the GDF model, a one-sided test of either $f_1 > 0$ (under the CE model) or of $f_2 > 0$ (under the AE model) is a valid test for family aggregation. A one-sided test of $f_2 > 0$ in the ACDE model that also includes f_1 is a valid test for genetic heritability. A two-sided test of $f_1 = 0$ in the ACDE model that includes f_2 is a valid test for genetic complexity. In the full ACDE model, the conditions $f_2 = f_1(\gamma_{MZ} - \gamma_{DZ})$ and $f_1 = -0.5 f_2(\gamma_{DZ} - 0)$ are equivalent to the triangle constraints. One-sided goodness-of-fit tests for violations of these conditions can be implemented, as can a test for $f_2 < 0$ ($\gamma_{MZ} < 0$). Tests and confidence intervals are based on the t-statistic, using standard errors and degrees of freedom described below.

Family resemblance

Equation (2) suggests a more general interpretation of γ_R . Suppose co-twin j 's residual

phenotype is $y_j - \sum_{k=0}^P b_k x_{jk}$. Then, twin i 's expected phenotype is increased by $\gamma_R \times \left(y_j - \sum_{k=0}^P b_k x_{jk} \right)$ after adjustment for covariates. That is, γ_R is the expected proportion of co-twin j 's residual phenotype shared by twin i . Under this second interpretation, we call γ_R the "family resemblance". For a binary trait, the resemblance is the added risk to twin i when co-twin j is affected (i.e. the recurrence risk after covariate adjustment).

GDF estimation

We use dual or double data entry as described for the original DF regression model. The data for each twin pair appears twice in the data set, once with twin i predicted by twin j and a second time with twin j predicted by twin i . Equation (2) is over-determined and does not yield a unique solution if f_1 , f_2 and the b_k s are all unknown. Instead, we use an iterative algorithm that loops over Steps 1 and 2 below, repeating the two steps multiple times until the algorithm converges.

Step 1 We estimate the fixed effects, i.e. the b_k s, using weighted regression for equation (3), $y_i = \sum_{k=0}^P b_k x_{ik} + e_i$. The inverse variance weights are $1/(1 + \gamma_R)$. With respect to the b_k s, uncorrelated pairs are equivalent to two independent observations, while perfectly correlated pairs are equivalent to a single observation. If $\hat{f}_1$ and $\hat{f}_2$ are the latest estimates of f_1 and f_2 from Step 2, then $\gamma_R = \hat{f}_1 + R \hat{f}_2$. Weights are set to one for all individuals the first time Step 1 is implemented.

Step 2 We estimate the DF effects, f_1 and f_2 , using a weighted regression for a model based on equation (2b). For each twin, we compute the phenotype residual $\hat{r}_i = y_i - \sum_{k=0}^P \hat{b}_k x_{ik}$ based on the latest estimates from Step 1. Then, equation (2b) can be rewritten as

$$\hat{r}_i = f_1 \hat{r}_j + f_2 R \hat{r}_j + e_i, \quad (4)$$

Equation (4) is solved using weighted regression with weights equal to $1/(1-\gamma_R^2)$, based on the latest values of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ from the previous execution of Step 2. Weights are set to one for all individuals the first time Step 2 is implemented.

Additional computational details for GDF

The two steps of the GDF algorithm are repeated until convergence is reached when the total absolute change for $\hat{\gamma}_1$ and $\hat{\gamma}_2$ from one iteration to the next is less than a user-specified tolerance value. We use a default tolerance of 10^{-6} . Degrees of freedom are computed as cluster-adjusted degrees of freedom for equation (3) minus 2 (for the ACDE model) or 1 (for the CE or AE model). One of two alternative methods is used to estimate standard errors that take account of the family structure of the data. The first approach is to use “sandwich” or “Huber-White” estimators of standard error with family treated as a clustering variable. (See Kohler and Rodgers 2001 for a detailed treatment.) This variation of the sandwich estimator provides standard errors that are robust to non-Normality and non-independence. The second approach is to compute bootstrap standard errors, based on random data sets created by resampling families (e.g. Angst et al. 2012^a; Angst et al. 2012^b; Ruau et al. 2012). This more computationally intensive approach requires generating several hundred random data sets and running the GDF algorithm on each. If data does not follow the Normal distribution, the linear regression model in Steps 1 and 2 can be replaced with a generalized linear model (GLM) with identity link. For the GLM version, tests and confidence intervals are based on the z-statistic. Although standard errors based on either the sandwich estimator or the bootstrap are robust to non-Normality, a GLM may improve fit when there is a reasonable choice of distribution.

Although not pursued in this paper, DF methods can be adapted to non-twin relationships and larger family structures as shown previously (Rodgers et al. 1994, Rodgers et al. 2001). In general, R is the expected identity by descent proportion (i.e. $1/2^d$, for d^{th} degree relatives). For genetically unrelated individuals, such as adoptive siblings, $R=0$. Different degrees of environmental sharing, such as occur in studies of twins raised apart, can also be accommodated using an analogous variable representing “environmental relatedness”. Additional types of relative pairs can resolve the unidentifiability of the ACDE model when the data include only twins raised together. When more than two individuals are present from the same family, a possible approach would be to enter each possible pair in the data set separately. Weights and resampling methods could potentially still yield efficient estimators and correct significance levels.

Mixed Effects Models

Two other estimation methods that can include covariates and are used for twin studies (Martin and Eaves 1977) are mixed effects maximum likelihood (MEML; Laird and Ware 1982) and structural equation models (SEM; Wright 1921). Inclusion of covariates in a model can improve efficiency (Schafer 1987; Stefanski and Carroll 1985) and focus the analysis on family correlations due to yet-to-be-discovered genetic and environmental factors. A variety of SEM and MEML twin software is available (e.g. Neale and Cardon, 1992; Williams et al. 1992; Neale 1997; Purcell 2002; Rijdsdijk and Sham 2002; Rabe-Hesketh and Skrondal 2008; Feng et al. 2009). In this paper, we focus on MEML as an alternative to the GDF algorithm.

MEML fits a two-level mixed effects model, with zygote random effects nested within family random effects. MEML usually assumes Normally distributed random errors and random effects, although probit and logit mixed effect twin models have been implemented (Hannah et al. 1985; MacGregor et al. 2000). Empirical studies have shown that variance-

component likelihood-ratio tests and model selection procedures can perform poorly if the error distribution is misspecified (Allison et al. 1999, Sullivan and Eaves 2002). The fact that the random-effect variances are non-negative automatically constrains MEML estimates to be consistent with the ACDE triangle. Therefore, MEML is subject to the limitations of constrained estimators discussed previously: including non-convergence, potentially biased estimates with skewed sampling distributions, unrepresentative standard errors and a lack of reliable confidence intervals. When a null hypothesis lies on the boundary of the parameter space, a modification of the chi-squared distribution is used for likelihood ratio tests (Neale 1997; Keller and Coventry 2005; Self and Liang, 1987; Stoel et al. 2006; Visscher 2006). The possibility of non-convergence means that MEML may be fit by a, potentially biased, adaptive strategy in which simpler models replace initial, more complicated models that fail to reach convergence.

SIMULATION DESIGN AND METHODS

Simulations to evaluate GDF and MEML were implemented in Stata. The full-factorial design included 3 factors:

- Sample size: $N(\text{pairs}) = 100$ or 400
- Error distribution: Normal(0,1) or exponential
- Twin model: One of 5 models given by $(\gamma_{DZ}, \gamma_{MZ})$
 - E, the null model (0, 0).
 - CE.1, a weak environmental model (0.1, 0.1)
 - CE.5, a strong environmental model (0.5, 0.5)
 - AE.2, a weak genetic model (0.1, 0.2)
 - AE.6, a strong genetic model (0.3, 0.6)

We generated 500 simulated data sets for each of the 20 possible factor combinations.

Each data set had equal numbers of MZ and DZ twin pairs. Two covariates were included. Age was generated as a discrete uniform random variable on the range 21–65. A second covariate, x was generated with mean zero, variance 1 and within-twin-pair correlation ρ . The value of ρ varied from data set to data set and was generated as a random uniform variable on the range $-0.75, 0.75$. By design, the expected trait value for each twin was $10 + 0.1 \text{ age} + 1.0 x$. For exponential error, within-pair covariance was created by letting the error term for each twin be the sum of two exponential variables, the first of which was shared by both members of a pair. The exponential term was centered and scaled to have the same mean and variance as the standard Normal distribution.

GDF and MEML were separately applied to each of the 20×500 random data sets to carry out hypothesis tests, estimate parameters and compute confidence intervals. Nominal significance levels were set at 5%. Nominal confidence levels were set at 95%. Each method included *age* and *x* as covariates in the model and assumed a Normal distribution, even for data sets with exponential error. Estimates and confidence intervals were evaluated in terms of both the GDF coefficients (f_1 and f_2) and the family resemblances (γ_{MZ} and γ_{DZ}), which were calculated as linear combinations of each other.

MEML simulation details

MEML was implemented using methods previous described (Angst et al. 2010), following a parameterization similar to Rabe-Hesketh and Skrondal (2008). Likelihood ratio tests were

used, adjusting the p -value when the null hypothesis is on a parameter boundary (Self and Liang 1987). Confidence intervals are not widely reported in the twin literature. We used the “delta” method for the log of the variance components as implemented in Stata (Stata User’s Guide 2009, pp. 283–284) to compute intervals for MEML, using the degrees freedom used for GDF. Performance rates reported for MEML tests, estimators and intervals included only data sets in which MEML reached convergence.

GDF simulation details

GDF was implemented as described above using Huber-White standard errors as implemented in Stata (Stata User’s Guide 2009, pp. 297–305), with tests and confidence intervals based on the appropriate t -statistic. Comparisons of GDF and MEML estimators are necessarily artificial as their parameter spaces differ. When the true values of γ_R are outside the ACDE triangle, MEML cannot correctly estimate them, making GDF the better approach. One could argue that, when the triangle constraints are known to be correct as in the simulation, it would be fair to compare MEML to a constrained GDF estimator. Thus, we considered a second GDF estimator to which triangle constraints were applied *post hoc*. Constrained GDF estimates were obtained by applying the following to the unconstrained GDF estimate.

1. If $\gamma_{MZ} < \gamma_{DZ}$, both were set to $\frac{\gamma_{MZ} + \gamma_{DZ}}{2}$.
2. If either $\gamma_{MZ} < 0$ or $\gamma_{DZ} < 0$, it was set to 0.
3. Corresponding constrained estimates of f_1 and f_2 were calculated as linear combinations of the constrained estimates of γ_{MZ} and γ_{DZ} from Step 2.

Standard errors and confidence intervals are not available for triangle-constrained GDF estimates.

RESULTS

Algorithm speed and convergence

The GDF ACDE model converged in 2 to 8 iterations in all simulated data sets. Generating one batch of 500 random data sets and fitting the GDF AE, CE and ACDE models required to implement the tests to each took at most 67 seconds. MEML failed to converge in 4.2% of data sets. Convergence failure ranged from 0% for AE.6 models with large N to 16% for E models with small N . Time to complete MEML for 500 data sets ranged from 12 to 205 minutes, with large N , null models taking the most time. Time per data set varied for MEML, with a small number of data sets consuming most of the time.

Hypothesis Test Performance

Both methods performed similarly in the E vs. AE test of family aggregation, maintaining good false positive rates (FPRs) on average and comparable power (Table 1). No FPR differed significantly from 5%, suggesting that Monte Carlo error from the simulation can explain the small difference observed. Results for the E vs. CE test of family aggregation (not shown) were remarkably similar to the E vs. AE test, doing slightly better under twin models CE.1 and CE.5 and slightly worse under AE.2 and AE.6. For the CE vs. ACDE genetic test, MEML FPRs were sometimes conservative, while GDF FPRs were slightly anti-conservative. However, MEML had particularly high FPRs under exponential error with a strong environmental effect, while GDF FPRs were less high. Larger environmental effects tended to increase the MEML FPR, but not the GDF FPR. Smaller samples sizes tended to increase the GDF FPR, but not the MEML FPR. For small N , GDF had better average power than MEML, 34.9% vs. 27.5%. For large N , the two methods had similar

power, 61.8% vs. 60.1%, respectively. Comparisons of power should consider the occurrence of some high FPRs for both methods. Data sets for which MEML did not converge were omitted from this analysis.

Bias

MEML produced biased estimates when the true model was near the boundaries of the ACDE triangle (Table 2). MEML bias was consistently increased in smaller samples, suggesting that small studies may contribute to the problem of over-estimated heritability. Bias was also greater for the exponential distribution than it was for the Normal distribution. The greatest bias was seen for f_2 , which is twice the difference of the MZ and DZ correlations. This is likely to translate into upwardly biased heritability estimates. Bias for f_1 tended to be negative, corresponding to over-estimation of D and under-estimation of C. Greater downward bias for f_1 occurred in the same settings as greater upward bias for f_2 . GDF estimates were essentially unbiased under all circumstances. Bias for the *post hoc* constrained GDF estimator (not shown) was very similar to bias for MEML. The absolute bias for the two constrained methods usually differed by no more than 0.02. For 7 of 40 possible estimates, MEML absolute bias was from 0.03 to 0.07 greater than that of the constrained GDF estimator.

Root mean squared error (MSE)

The trade-off for the negligible bias of the unconstrained GDF estimators was substantially more variability in settings where MEML's triangle constraints were known to be true (data not shown). To compare MEML and the constrained GDF estimator, we calculated the ratio of the root mean squared errors (RRMSE) for the two methods (Table 3). $RRMSE > 1$ favored MEML. $RRMSE < 1$ favored the constrained GDF estimator. Neither method dominated on this measure. GDF did better 32 of 48 times under the null or weak E, CE.1 and AE.2 models, while MEML did better 24 of 32 times under the strong CE.5 and AE.6 models. Proximity to the constraint boundaries affects the relative efficiency of MEML compared to the constrained GDF estimator. We recommend use of the unconstrained GDF estimator. However, the similarity of the MEML estimator and the constrained GDF estimator if similar constraints are applied to both is interesting.

Confidence intervals

GDF confidence interval coverage for f_1 , f_2 , γ_{MZ} and γ_{DZ} (Table 3) was on average 94.6% for large N (93.2%–96.0%) and 91.5% (88.0%–94.4%) for small N . GDF coverage was on average 1% worse for the exponential model than for the Normal model, but was similar across ACDE models. It is possible that bootstrap standard errors would improve coverage for small samples relative to the Huber-White standard errors used in the simulation. MEML coverage (not shown) averaged 74.0% (range 31.4%–95.0%) for large N and 67.7% (range 27.8%–95.2%) for small N . MEML coverage was worst for the E model and best for the AE.6 model. MEML coverage for exponential error was 5% worse on average than for Normal error. Contributing factors to poor MEML coverage near the ACDE boundaries were problems with standard error estimates and asymmetric sampling distributions. We therefore do not recommend that these particular MEML confidence intervals be adopted. Bootstrap standard errors are not practical for MEML for reasons related to the constrained parameter space.

An example: Bronchopulmonary dysplasia

Bronchopulmonary dysplasia (BPD) is a severe chronic lung disorder in premature infants associated with a variety of lifelong developmental impairments. We applied GDF to a multicenter, retrospective twin study of infants between 24 and 32 weeks of gestational age

(GA). Bhandari et al. (2006) estimated the heritability of BPD in these data to be 53% after adjustment for covariates, including a linear effect of GA. Twins with unusually high birth weight (>3000 grams) or missing data were removed from the analysis, leaving 396 babies from 137 DZ and 62 MZ twin pairs. The BPD rate was 24% (94/396) and ranged from 70% (7/10) at 24 weeks GA to 7% (6/83) at 32 weeks GA. We used the GLM version of GDF, specifying a binomial distribution with an identity link and offset. To simplify the example, we included main and quadratic effects GA, but no other covariates.

The AE family test detected significant family aggregation (one-sided $p < 0.001$). The genetic effect in the ACDE model was significant (one-sided $p = 0.006$). See Figure 3 for the estimates from this model. The non-significance of f_1 suggests that BPD may be a simple additive, genetic phenotype except for the effect of GA. Thus, genetic interactions and shared family environment may have little importance in its etiology. BPD risk decreased substantially with increased GA. Under the ACDE model, the fitted BPD risk was 76% for a baby born at 24 weeks and 8% for a baby born at 32 weeks. Under the family resemblance interpretation, GDF provides covariate-adjusted estimates of the increased risk to a twin implied by his or her co-twin's BPD diagnosis. According to the estimated model, a BPD-positive MZ co-twin increases the risk for a 24-week baby to 93%, while a BPD-negative MZ co-twin decreases the risk to 23%. For a DZ twin pair, these risks are 83% and 53%, respectively. Co-twins with an unexpected outcome (e.g. no BPD at 24 weeks GA) alter the predicted risk more than co-twins whose outcome was expected. MZ twins provide more information than DZ twins with respect to a genetic trait such as BPD appears to be. Although f_1 was not significant, the estimated ACDE model was in the ADE region ($f_1 < 0$). Thus, f_2 in the ACDE model could not be interpreted as heritability estimate without further assumptions. We assumed that $D = 0$ and refit the AE model without f_1 (effectively setting f_1 and D to 0) and estimated the GA-adjusted heritability as 70% (CI: 52%–89%). When $f_1 < 0$, the best-fitting AE model and the best-fitting ACE model are identical.

Results were similar if GDF with least-squares regression was used instead of the GLM version. In comparison to the GLM estimates shown in Figure 3, least-squares GDF estimates were $\gamma_{DZ} = 0.30$ (95% CI: 0.09–0.49) and $\gamma_{MZ} = 0.72$ (95% CI: 0.52–0.93). When GA was omitted from the GLM analysis, the estimate of γ_{DZ} increased and the genetic effect was not significant ($p = 0.08$).

DISCUSSION

We have introduced a generalized DeFries-Fulker regression model for the analysis of twin data that provides a useful alternative to existing methods. Unlike the original model, the proposed model shares with MEML and SEM the ability to simultaneously estimate family effects due to latent genetic and shared environmental effects and fixed effects due to observed covariates. This feature is important for understanding the relationships among the factors underlying complex traits. GDF estimates and tests are competitive with MEML methods. However, the GDF approach enables a broader view of twin analysis. We recommend the unconstrained version of GDF, because it can provide unbiased estimates, valid goodness-of-fit tests and reliable confidence intervals. However, if one chooses to accept the triangle constraints *a priori*, the constrained version of the GDF estimator performs similarly to MEML, but has superior computational properties. In the settings we have examined, root MSE of constrained GDF estimates ranges from 20% better to 10% worse than MEML, while bias is slightly better.

Although not as simple as the original DF regression method, the GDF model has advantages over maximum likelihood methods. By building on DF's computational simplicity, GDF has more freedom to accommodate more complicated models. It may be

possible for GDF to address different covariate models (e.g. splines or polynomials); error distributions; and data types (e.g. censored survival time data or rating scores with ceiling effects). It may also be possible to extend the GDF model to explore gene by environment (G×E) interactions. For example, the model for BPD risk in our example could be expanded to include interactions of GA with $(y_j - \sum_{k=0}^p b_k x_{jk})$ and $R \times (y_j - \sum_{k=0}^p b_k x_{jk})$, which would allow the heritability to depend on GA. The same strategy would be useful for age-of-onset in studies of prostate cancer or Alzheimer's disease. The G×E model would have the potential to identify separate, optimal sub-populations for future genetic association studies. Other possible extensions include larger families, non-twin relationships and adoption studies.

Neither GDF nor MEML dominated in every situation with respect to test and estimator performance. However, an important advantage of GDF regression is the availability of working confidence intervals. While coverage was approximately 3% below nominal levels in small samples, coverage was on target for large samples. Bootstrap standard errors or confidence intervals may improve coverage for small samples over that seen in the simulation. In contrast, our simulations demonstrate the unreliability of delta-method confidence intervals for MEML. Common practice in twin studies has been to report *p*-values, but not standard errors or confidence intervals. We strongly encourage the reporting of measures of uncertainty for heritability and other estimates. Furthermore, we believe that studies should focus on the MZ and DZ correlations or similarities and always report estimates and confidence intervals for these parameters. Estimates of heritability under the ACDE model necessarily require additional subjective judgments, potentially creating a distorted impression. Reporting correlations would provide greater transparency and comparability of results across twin studies.

The GDF model also yields a useful, unified interpretation of twin data in terms of the family resemblance. This interpretation ties together the correlation used in quantitative trait analysis with the recurrence risk used for binary outcomes. Furthermore, it describes the relationship between twin phenotypes in a way that applies to almost any data distribution through its expectation and translates directly into prediction models.

Limitations of GDF regression include the anti-conservative false positive and confidence interval coverage rates observed in the simulations, especially for smaller samples. Small sample performance requires further work. The simulation addressed only a limited set of circumstances, which may not be representative. Simulation results may not apply in other settings, to other implementations of MEML or to SEM. Different implementations of MEML might also have yielded somewhat different results. As proposed, GDF does not take advantage of the greater information provided by larger, extended families.

The current debate regarding “missing heritability” may in fact reflect misestimated heritability. The GDF strategy introduced here includes new and more extensible models for twin data; reduced reliance on unverifiable assumptions and the constraints of the ACDE model; less focus on estimation of unidentifiable parameters; improved confidence intervals as measures of uncertainty; and the potential for new goodness-of-fit tests. Taken together, these advantages may help more accurately determine how much heritability there is to explain in the first place. We believe it is also important to revisit existing twin or family samples with the results of genetic association studies to better assess the extent of the missing heritability problem. Thus, twin studies can complement genetic association studies by developing the evidence for their design beforehand and assessing their findings after the fact. The GDF model will facilitate such studies by providing an alternative set of flexible and robust analytic tools.

Acknowledgments

The authors thank Drs. Heping Zhang and Vineet Bhandari for access to the BPD data. We also thank Joe Rodgers and an anonymous reviewer for many helpful suggestions. This work was supported by NIH grants R01MH086135, R01DA023063 and R01MH083972 and the Clinical Science Research & Development Service of the Department of Veterans Affairs (A Cooperative Studies Program-Wide DNA Bank, CSP#478).

References

- Allison DB, Neale MC, Zannolli R, Schork NJ, Amos CI, Blangero J. Testing the Robustness of the Likelihood-Ratio Test in a Variance-Component Quantitative-Trait Loci-Mapping Procedure. *American Journal of Human Genetics*. 1999; 65:531–544. [PubMed: 10417295]
- Angst MS, Lazzeroni LC, Phillips NG, Drover DR, Tingle M, Ray A, Swan GE, Clark JD. Aversive and reinforcing opioid effects: a pharmacogenetic twin study. *Anesthesiology*. 2012; 117 (1):22–37. [PubMed: 22713632]
- Angst MS, Phillips NG, Drover DR, Tingle M, Ray A, Swan GE, Lazzeroni LC, Clark JK. Pain sensitivity and opioid analgesia: a pharmacogenetic twin study. *Pain*. 2012; 153(7):1397–409. [PubMed: 22444188]
- Angst MS, Phillips NG, Drover DR, Tingle M, Galinkin JL, Christians U, Swan GE, Lazzeroni LC, Clark JD. Opioid Pharmacogenomics Using a Twin Study Paradigm: Methods and Procedures for Determining Familial Aggregation and Heritability. *Twin Research and Human Genetics*. 2010; 13(5):412–425. [PubMed: 20874462]
- Bhandari V, Bizzaro MJ, Shetty A, Zhong X, Grier PP, Zhang H, Ment LR. Familial and Genetic Susceptibility to Major Neonatal Morbidities in Preterm Twins. *Pediatrics*. 2006; 117(6):1901–1906. [PubMed: 16740829]
- Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. *Nature Reviews Genetics*. 2002; 3:872–882.
- Carey G. Cholesky Problems. *Behavior Genetics*. 2005; 35(5):653–665. [PubMed: 16184491]
- Castillo E, Galambos J. Conditional Distributions and the bivariate normal distribution. *Metrika*. 1989; 36(1):209–214.
- Cherny SS, DeFries JC, Fulker DW. Multiple Regression Analysis of Twin Data: A Model-Fitting Approach. *Behavior Genetics*. 1992; 22(4):489–497. [PubMed: 1503550]
- DeFries JC, Fulker DW. Multiple regression analysis of twin data. *Behavior Genetics*. 1985; 15(5):467–473. [PubMed: 4074272]
- Feng R, Zhou G, Zhang H. Analysis of twin data using SAS. *Biometrics*. 2009; 65(2):584–589. [PubMed: 18647295]
- Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. *Philosophical Transactions of the Royal Society of Edinburgh*. 1918; 52:399–433.
- Friedman MC, Chhabildas N, Budhiraja N, Willcutt EG, Pennington BF. Etiology of the comorbidity between RD and ADHD: Exploration of the non-random mating hypothesis. *American Journal of Medical Genetics*. 2003; 120B(1):109–115. [PubMed: 12815748]
- Galton F. The history of twins, as a criterion of the relative powers of nature and nurture. *Fraser's Magazine*. 1875; 12:566–576.
- Hannah MC, Hopper J, Mathews J. Twin concordance for a binary trait. II. Nested analysis of ever-smoking and ex-smoking traits and unnested analysis of a “committed-smoking” trait. *American Journal of Human Genetics*. 1985; 37(1):153–165. [PubMed: 4038848]
- Haseman JK, Elston RC. The estimation of genetic variance from twin data. *Behavior Genetics*. 1970; 1(1):11–19. [PubMed: 5535855]
- Hettema JM, Neale MC, Kendler KS. Physical similarity and the equal-environment assumption in twin studies of psychiatric disorders. *Behavior Genetics*. 1995; 25(4):327–335. [PubMed: 7575361]
- Hicks BM, Krueger RF, Iacono WG, McGue M, Patrick CJ. Family Transmission and Heritability of Externalizing Disorders: A Twin-family Study. *Archives of General Psychiatry*. 2004; 61(9):922–928. [PubMed: 15351771]

- Karlin S, Cameron EC, Chakraborty R. Path Analysis in Genetic Epidemiology: A Critique. *American Journal of Human Genetics*. 1983; 35:695–732. [PubMed: 6349335]
- Keller MC, Coventry WL. Quantifying and addressing parameter indeterminacy in the classical twin design. *Twin Research and Human Genetics*. 2005; 8:201–213. [PubMed: 15989748]
- Kemphorne O, Osborne RH. The Interpretation of Twin Data. *American Journal of Human Genetics*. 1961; 13(3):320–339. [PubMed: 13752449]
- Kendler KS. A test of the equal environment assumption in twin studies of psychiatric illness. *Behavior Genetics*. 1993; 23:21–27. [PubMed: 8476388]
- Kohler H, Rodgers JL. DF-like Analyses of Binary, Ordered, and Censored Variables Using Probit and Tobit Approaches. *Behavior Genetics*. 1999; 29(4):221–232.
- Kohler H, Rodgers JL. DF-Analyses of Heritability with Double-Entry Twin Data: Asymptotic Standard Errors and Efficient Estimation. *Behavior Genetics*. 2001; 31(2):179–191. [PubMed: 11545535]
- Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982; 38:963–974. [PubMed: 7168798]
- MacGregor AJ, Antoniadou L, Matson M, Andrew T, Spector TD. The genetic contribution to radiographic hip osteoarthritis in women: Results of a classic twin study. *Arthritis & Rheumatism*. 2000; 43(11):2410–2416. [PubMed: 11083262]
- Manolio T, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, et al. Finding the missing heritability of complex diseases. *Nature*. 2009; 461:747–753. [PubMed: 19812666]
- Martin NG, Eaves LJ. The genetical analysis of covariance structure. *Heredity*. 1977; 38:79–95. [PubMed: 268313]
- Martin NG, Eaves LJ, Heath AC, Jardine R, Feingold LM, Eysenck HJ. Transmission of social attitudes. *Proc Natl Acad Sci USA*. 1986; 83:4364–4368. [PubMed: 3459179]
- Neale, M.; Cardon, L. *Methodology for genetic studies of twins and families*. Kluwer Academic Publishers; 1992.
- Neale, MC. *Mx: Statistical modeling*. 2. Richmond: Medical College of Virginia, Department of Psychiatry; 1997.
- Purcell S. Variance Components Models for Gene-Environment Interaction in Twin Analysis. *Twin Research*. 2002; 5(6):554–571. [PubMed: 12573187]
- Purcell S, Sham PC. A Model-Fitting Implementation of the DeFries-Fulker Model for Selected Twin Data. *Behavior Genetics*. 2003; 33:271–278. [PubMed: 12837017]
- Rabe-Hesketh S, Skrondal A, Gjessing HK. Biometrical Modeling of Twin and Family Data Using Standard Mixed Model Software. *Biometrics*. 2008; 64(1):280–288. [PubMed: 17484777]
- Rijsdijk FV, Sham PC. Analytic approaches to twin data using structural equation models. *BRIEFINGS IN BIOINFORMATICS*. 2002; 3(2):119–133. [PubMed: 12139432]
- Risch N. Linkage Strategies for Genetically Complex Traits. II. The Power of Affected Relative Pairs. *American Journal of Human Genetics*. 1990; 46:229–241. [PubMed: 2301393]
- Rodgers JL, Buster M, Rowe DC. Genetic and environmental influences on delinquency: DF analysis of NLSY kinship data. *Journal of Quantitative Criminology*. 2001; 17:145–168.
- Rodgers JL, Kohler HP. Reformulating and Simplifying the DF Analysis Model. *Behavior Genetics*. 2005; 35(2):211–217. [PubMed: 15685433]
- Rodgers JL, McGue M. A Simple Algebraic Demonstration of the Validity of DeFries-Fulker Analysis in Unselected Samples with Multiple Kinship Levels. *Behavior Genetics*. 1994; 24(3):259–262. [PubMed: 7945155]
- Rodgers JL, Rowe DC, May K. DF analysis of NLSY IQ/achievement data: Nonshared environmental influences. *Intelligence*. 1994; 19:157–177.
- Ruau D, Dudley JT, Chen R, Phillips NG, Swan GE, Lazzeroni LC, Clark JD, Butte AJ, Angst MS. *PLoS Comput Biol*. 2012; 8(6):e1002538. [PubMed: 22685391]
- Schafer DW. Covariate measurement error in generalized linear models. *Biometrika*. 1987; 74(2):385–391.

- Self SG, Liang KY. Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of the American Statistical Association*. 1987; 82(398):605–610.
- Sham PC, Walters EE, Neale M, Heath AC, MacLean CJ, Kendler KS. Logistic regression analysis of twin data: Estimation of parameters of the multifactorial liability-threshold model. *Behavior Genetics*. 1994; 24:229–238. [PubMed: 7945153]
- StataCorp. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP; 2009.
- Stata Press. Stata User's Guide: Release 11. College Station, TX: StataCorp LP; 2009.
- Stefanski LA, Carroll RJ. Covariate measurement error in logistic regression. *Annals of Statistics*. 1985; 13:1335–1351.
- Stoel R, Garre FG, Dolan C, Wittenboer GV. On the Likelihood Ratio Test in Structural Equation Modeling When Parameters Are Subject to Boundary Constraints. *Psychological Methods*. 2006; 11(4):439–455. [PubMed: 17154756]
- Sullivan PF, Eaves LJ. Evaluation of Analyses of Univariate Discrete Twin Data. *Behavior Genetics*. 2002; 32(3):221–227. [PubMed: 12141783]
- Visscher PM. A Note on the Asymptotic Distribution of Likelihood Ratio Tests to Test Variance Components. *Twin Research and Human Genetics*. 2006; 9(4):490–495. [PubMed: 16899155]
- Visscher PM, Hill WG, Wray NR. Heritability in the genomics era—concepts and misconceptions. *Nature Rev Genet*. 2008; 9:255–266. [PubMed: 18319743]
- Waller NB. A DeFries and Fulker regression model for genetic nonadditivity. *Behavior Genetics*. 1994; 24(2):149–153. [PubMed: 8024531]
- Williams CJ. On the covariance between parameter estimates in models of twin data. *Biometrics*. 1993; 49:557–568. [PubMed: 8369388]
- Williams CJ, Christain JC, Norton JA. TWINAN90: A FORTRAN program for conducting ANOVA-based and likelihood-based analyses of twin data. *Computer Methods and Programs in Biomedicine*. 1992; 38:167–176. [PubMed: 1458866]
- Wright S. Correlation and causation. *Journal of Agricultural Research*. 1921; 20:557–585.
- Young SE, Stallings MC, Corley RP, Krauter KS, Hewitt JK. Genetic and environmental influences on behavioral disinhibition. *American Journal of Medical Genetics*. 2000; 96(5):684–695. [PubMed: 11054778]

Figure 1. The ACDE triangle

The horizontal and vertical axes are constraints that are part of the classical ACDE model, which assumes non-negativity of the variance components.

Figure 2.
Alternative interpretations of $\gamma_{MZ} > 2 \gamma_{DZ}$.

```
. twdf BPD GestAge GestAgeSq, t(ACDE) vce(cluster family)
```

```
Numbers of obs (eq1, eq2) = 396, 394    Degrees of freedom = .
Converged after 5 iterations at 1.000e-06 tolerance
```

	BPD	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
fe	GestAge	-.5727566	.2538489	-2.26	0.024	-1.070291	-.0752219
	GestAgeSq	.0087255	.0043912	1.99	0.047	.0001189	.0173321
	_cons	9.476174	3.650439	2.60	0.009	2.321446	16.6309
df	f1	-.1047707	.2522843	-0.42	0.678	-.5992388	.3896973
	f2	.8006093	.3148374	2.54	0.011	.1835394	1.417679
zy	GammaDZ	.2955339	.1138225	2.60	0.009	.0724459	.518622
	GammaMZ	.6958386	.1087432	6.40	0.000	.4827058	.9089714
	DiffMZ_DZ	.4003047	.1574187	2.54	0.011	.0917697	.7088397

Figure 3. Screen snapshot of BPD analysis

Command line and output from Stata results window for the full two-parameter GDF model of BPD risk, using a binomial GLM, with quadratic effect of gestational age. Displayed *p*-values are two-sided.

Table 1

Test Performance

Model	Family Test ¹		Genetic Test ²	
	GDF	MEML	GDF	MEML
E	FPR³ (%)			
100 Normal	6.0	4.4	6.0	1.0
100 Exponential	3.4	5.0	7.8	2.6
400 Normal	5.6	5.2	5.2	3.0
400 Exponential	3.6	4.8	4.4	2.6
CE.1	Power (%)		FPR³ (%)	
100 Normal	23.8	22.0	5.2	1.2
100 Exponential	20.2	21.6	6.8	5.2
400 Normal	56.6	55.4	6.0	3.0
400 Exponential	59.2	56.8	4.8	5.2
CE.5	Power (%)		FPR³ (%)	
100 Normal	100.0	99.4	4.8	1.8
100 Exponential	99.2	99.6	8.0	13.0
400 Normal	100.0	99.2	5.2	2.0
400 Exponential	100.0	99.0	6.2	14.0
AE.2	Power (%)			
100 Normal	48.0	44.8	13.2	5.0
100 Exponential	42.6	41.8	16.0	8.2
400 Normal	93.8	90.8	30.0	15.2
400 Exponential	91.4	88.8	26.6	30.0
AE.6	Power (%)			
100 Normal	100.0	98.6	61.8	49.0
100 Exponential	99.8	99.2	48.4	47.6
400 Normal	100.0	100.0	98.8	98.4
400 Exponential	100.0	100.0	91.6	96.8

¹ ACDE vs. CE test.

² AE vs. E test. Results were similar for CE vs. E test (not shown).

³ FPR = false positive rate.

Table 2

Bias

Model	GDF				MEML			
	γ_{yz}	γ_{Mz}	f_1	f_2	γ_{yz}	γ_{Mz}	f_1	f_2
E								
100 Normal	-0.01	0.00	-0.02	0.02	0.03	0.07	-0.01	0.08
100 Exponential	-0.01	-0.01	-0.02	0.01	0.03	0.07	-0.02	0.09
400 Normal	-0.01	0.00	-0.01	0.01	0.01	0.04	-0.01	0.05
400 Exponential	0.00	0.00	-0.01	0.01	0.01	0.04	-0.01	0.05
CE.1								
100 Normal	-0.01	0.00	-0.02	0.02	-0.01	0.04	-0.07	0.12
100 Exponential	-0.01	-0.01	-0.01	0.01	-0.02	0.05	-0.10	0.15
400 Normal	-0.01	0.00	-0.01	0.01	-0.02	0.02	-0.06	0.08
400 Exponential	0.00	0.00	-0.01	0.01	-0.02	0.03	-0.06	0.09
CE.5								
100 Normal	-0.01	-0.01	-0.02	0.01	-0.03	0.02	-0.09	0.11
100 Exponential	-0.02	-0.02	-0.03	0.02	-0.05	0.03	-0.14	0.17
400 Normal	-0.01	0.00	-0.01	0.00	-0.02	0.01	-0.05	0.06
400 Exponential	-0.01	0.00	-0.01	0.01	-0.02	0.02	-0.07	0.08
AE.2								
100 Normal	-0.01	-0.01	-0.02	0.01	0.00	0.02	-0.01	0.04
100 Exponential	-0.01	-0.01	-0.02	0.01	0.00	0.03	-0.02	0.06
400 Normal	-0.01	0.00	-0.01	0.01	-0.01	0.00	-0.02	0.02
400 Exponential	0.00	0.00	-0.01	0.01	0.00	0.01	-0.02	0.03
AE.6								
100 Normal	-0.01	-0.01	-0.02	0.00	-0.01	0.00	-0.01	0.00
100 Exponential	-0.02	-0.02	-0.02	0.00	-0.01	0.00	-0.01	0.01
400 Normal	-0.01	0.00	-0.01	0.00	-0.01	0.00	-0.01	0.01
400 Exponential	-0.01	0.00	-0.01	0.00	0.00	0.00	-0.01	0.01

Table 3

RRMSE and Coverage.

Model	RRMSE ¹ (MEML/GDF)				Coverage ² (GDF)			
	γ_{0z}	γ_{Mz}	$f1$	$f2$	γ_{0z}	γ_{Mz}	$f1$	$f2$
E								
100 Normal	0.96	0.91	1.04	1.04	90.2	93.8	92.2	93.6
100 Exponential	0.88	0.98	0.89	0.85	90.4	88.0	90.0	91.8
400 Normal	0.81	0.90	0.96	0.85	94.6	93.4	94.8	94.8
400 Exponential	0.83	0.91	0.95	0.87	93.2	94.0	94.0	96.0
CE.1								
100 Normal	0.96	1.01	1.07	1.03	90.4	93.2	92.2	93.2
100 Exponential	0.87	1.01	0.91	0.85	88.6	91.2	91.2	91.6
400 Normal	0.96	1.02	1.00	0.97	94.2	93.8	95.6	95.8
400 Exponential	0.90	1.01	0.97	0.91	93.2	93.4	94.8	95.2
CE.5								
100 Normal	1.00	1.06	1.11	1.11	92.6	92.6	93.2	94.4
100 Exponential	1.00	1.00	0.99	0.99	89.8	91.6	91.2	91.0
400 Normal	1.01	1.04	1.06	1.06	94.2	95.8	95.4	95.4
400 Exponential	1.00	1.02	1.03	1.03	94.6	94.6	94.4	94.4
AE.2								
100 Normal	0.97	1.00	1.03	1.01	90.6	93.0	92.2	93.6
100 Exponential	0.90	0.98	0.93	0.89	88.8	90.6	91.2	92.2
400 Normal	1.00	1.01	1.01	1.01	94.2	94.2	95.4	95.8
400 Exponential	0.95	1.00	0.97	0.95	93.2	93.4	94.8	94.4
AE.6								
100 Normal	1.06	1.04	1.07	1.08	91.0	93.2	91.6	93.4
100 Exponential	1.07	0.95	0.95	0.97	89.6	90.2	90.4	90.6
400 Normal	1.08	1.01	1.03	1.06	94.4	95.6	95.6	95.8
400 Exponential	1.06	0.96	0.98	1.00	94.6	94.4	95.0	95.0

¹RRMSE for MEML vs. the *post hoc* constrained GDF estimator.²Confidence interval coverage (%) coverage for the original GDF estimator