NIASRA

NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

The University of Wollongong

Working Paper

13-14

Using Social Network Information for Survey Estimation

Thomas Suesse and Ray Chambers

National Institute for Applied Statistics Research Australia, University of Wollongong,
Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845.
Email: anica@uow.edu.au

Using Social Network Information for Survey Estimation

Thomas Suesse

National Institute for Applied Statistics Research Australia

University of Wollongong, Wollongong, Australia

E-mail: tsuesse@uow.edu.au

Ray Chambers

National Institute for Applied Statistics Research Australia

University of Wollongong, Wollongong, Australia

E-mail: ray@uow.edu.au

Summary. Model-based and model-assisted methods of survey estimation aim to improve the precision of estimators of the population total or mean relative to methods based on the nonparametric Horvitz-Thompson estimator. These methods often use a linear regression model defined in terms of auxiliary variables whose values are assumed known for all population units. Information on networks represents another form of auxiliary information that might increase the precision of these estimators, particularly if it is reasonable to assume that networked population units have similar values of the survey variable. Linear models that use networks as a source of auxiliary information include autocorrelation, disturbance and contextual models. In this paper we focus on social networks, and investigate how much of the population structure of the network needs

to be known for estimation methods based on these models to be useful. In particular, we use simulation to compare the performance of the best linear unbiased predictor under a model that ignores the network with model-based estimators that incorporate network information. Our results show that incorporating network information via a contextual model seems to be the most appropriate approach. We also show that one does not need to know the full population network, but that knowledge of the partial network linking the sampled population units to the non-sampled population units is necessary.

Keywords: BLUP, social network models, linear models, model-based survey estimation

1 Introduction

Survey estimation typically focuses on estimating the total $T_Y = \sum_{i \in U} Y_i$ of the values of a variable Y defined over a finite population U. Here $i \in U$ denotes the N units making up the population U. Given a sample s of n units from U, T_Y is usually estimated by $\hat{T}_Y = \sum_{i \in s} w_i Y_i$, where the w_i are sample weights and $i \in s$ denotes the n units in the sample. Traditionally, these weights are expansion weights, i.e. w_i is the inverse of the selection probability of the ith population unit. However, expansion weights can be quite inefficient, and alternative weighting methods derived from model-based and model-assisted methods of survey estimation, see Chambers and Clark (2012) and Srndal et al. (1992), are used to increase the precision of $\hat{T}_Y$. In most cases this is done by defining the sample weights so that $\hat{T}_Y$ is an efficient unbiased predictor of T_Y under a linear regression model for Y in terms of a multivariate auxiliary variable X.

Population regression models that link an individual's value of Y to auxiliary variables corresponding to that individual's geographic location, gender and age are commonly used in survey estimation. However, auxiliary information can be more complex than this. In particular, information about other individuals in the population that are 'linked' to a particular individual also constitutes auxiliary information about that individual. This is sometimes referred to as $network$

information, and typically indicates between individual correlation in the population values of Y. In this paper we describe model-based survey estimation methods that exploit auxiliary information about population networks. In particular, we describe how the specification of the best linear unbiased predictor (BLUP) of T_Y can be tailored to allow for between individual correlation induced by the presence of a population network.

In order to motivate the use of network information in survey estimation, consider the case of the British Household Panel Study (BHPS, `http://www.iser.essex.ac.uk/survey/bhps/`). This is an annual longitudinal survey of British households that has been conducted since 1991. It is based on a sample of approximately $5,500$ households, covering more than $10,000$ individuals. The main objective of the survey is to further the understanding of social and economic change at the individual and household level in Britain. However, in addition to information about the surveyed individual, the BHPS also provides information about a person's three closest friends. Variables collected on the three closest friends are: age, sex, ethnicity, distance to friend (< 1 mile, between 1 and < 5 miles, between 5 and 50 miles, > 50 miles) and unemployment status. This information is available in seven waves, corresponding to the even-numbered years 1992 - 2004.

Because friends tend to share common characteristics, it is plausible that the BHPS information on friendship ties may be of value when modelling the other survey variables, in the same way as the ties between household members are typically viewed as influential in determining the outcomes of many social and economic variables. For example, a person whose friends are older than the norm might have a higher than average income, even after adjusting for that person's age and gender. As a consequence, one might think of also controlling for the average age of friends when predicting a person's income. A model of this type is referred to as a *contextual* model in what follows since it controls for contextual effects, such as the average age of friends. Clearly, since the BHPS collects information on a person's three best friends, there is scope for applying a contextual model when estimating using BHPS data. This might lead to more precise survey estimates, as a contextual variable represents an additional source of information.

The friendship data collected in the BHPS are a special case of a general type of auxiliary

data whose availability is becoming increasingly widespread, especially with the rapid uptake of modern telecommunications technology. This is network data, defined by the existence, direction and strength of relationships between individuals in a population of interest. Statistical modelling of networks is now reasonably well established, see for example Frank and Strauss (1986); Snijders (2002); Hunter and Handcock (2006), though applications to very large networks (e.g. defined by populations similar in size to those covered by a survey like the BHPS) are still rare, with data on very large networks now considered to be part of the ubiquitous Big Data concept. Furthermore, we are not aware of any attempt to use the information in a network defined on a population of interest to improve survey estimation for that population, although, as the argument put forward in the previous paragraph indicates, there may be value in doing so.

In order to use network information in a model linking a survey variable Y to an auxiliary variable X we need to characterise the population network as the outcome of a random process. In this context, we focus in this paper on a network that identifies the existence and direction of a relationship between individuals in a population of size N. It is standard to represent such a network by a matrix of zeros and ones, $\mathbf{Z} = (Z_{ij})_{i,j=1}^{N}$ with $Z_{ii} = 0$ by convention. If a relationship exists between two individuals i and j, then $Z_{ij} = 1$ and we refer to i and j as being linked; otherwise $Z_{ij} = 0$. Such a network is said to be undirected if $\mathbf{Z} = \mathbf{Z}'$, otherwise it is a directed network.

Networks are most useful when characteristics of the individuals that make up the population covered by the network are also known. In such networks one not only knows the characteristics of a particular individual, but also the characteristics of the other individuals in the population linked to that individual via the network. This $external$ auxiliary information may be useful in discriminating between individuals, and hence may be useful in prediction, the ultimate goal of survey estimation. For example, the BHPS collects information about the three best friends of a surveyed individual, without identifying the friends. Given that the links corresponding to being 'one of three best friends' define a network, this information can be treated as auxiliary data for the surveyed individual, and, combined with a model for the network, may help with formulating

a more efficient prediction model for the population.

Linear models that use a social network as additional information to model the expected value of a response variable include contextual network (CN) models (Friedkin 1990). However, this information can also be used to model between unit correlation in the population values of the response variable. Such second order models include network effects models, also known as autocorrelation (AR) models, and network disturbance (ND) models (Ord 1975; Doreian et al. 1984; Duke 1993; Marsden and Friedkin 1993; Leenders 2002).

When the network defined by $\mathbf{Z}$ is known for all N individuals in the population, the CN, AR and ND population models can be used for survey estimation. However, in practice it is extremely unlikely that $\mathbf{Z}$ will be fully known, and a more realistic scenario is one where one or more components of this matrix will be known. The most obvious is where only the component $\mathbf{Z}_{ss}$ corresponding to the sub-network of relationships between the n sampled individuals in s is known. Unless the sampling fraction is large, or the sample is highly clustered, it is unlikely that this sub-network will contain much useful information. Of more use, perhaps, is the component $\mathbf{Z}_{sr}$, defined by the links between the sampled individuals and the remaining $N-n$ non-sampled individuals in the population, denoted collectively by r. Clearly, if the network is an undirected one, the links from the non-sampled individuals to the sampled individuals will then also be known since, under symmetry, $\mathbf{Z}_{rs} = \mathbf{Z}'_{sr}$. The remaining component of $\mathbf{Z}$ is $\mathbf{Z}_{rr}$, which corresponds to the sub-network defined by the links between the $N-n$ non-sampled individuals in the population. This will generally be unknown. Using network information in a survey sampling context therefore implies that one has to deal with situations where partial network information is observed. This inevitably means that one needs to either use more complicated modelling methods or that one needs to somehow impute the missing network components.

The main focus of this paper is on the potential use of network information in survey estimation. In particular, we aim to address three questions: (i) Is embedding network information useful for survey estimation based on linear models? (ii) If the answer to (i) is yes, then which network models are potentially useful? and (iii) How much network data needs to be collected in order to

obtain potentially higher precision for survey estimation? In Section 2 we provide some context for these questions by defining a standard linear model that is often used for survey estimation, as well as its extension to a linear mixed model that allows for random cluster effects. Neither of these models incorporate network information, so we then describe three widely used linear models that allow for the availability of network information in addition to standard covariate and cluster information.

In Section 3 we briefly discuss estimation of the population mean of a survey variable using the empirical version of the BLUP (typically referred to as the empirical best linear unbiased predictor or EBLUP) based on a linear model for this variable, and its application under the network models introduced in the previous Section. In Section 4, the exponential random graph model (ERGM) for a network is introduced and its use in imputation of missing network information is described, with the aim of using this imputed information in the network model-based estimators introduced in Chapter 3. These ideas are then brought together in Section 5 where we describe a simulation study that investigates the performances of the imputation-based EBLUPs defined by these different network models. In particular, we compare these estimators with the standard linear estimators that ignore network information. Section 6 completes the paper with a discussion of our findings as they relate to the three questions raised above.

2 Linear Models on Networks

In this section we describe a number of population level linear models that use network information. Throughout, we use a friendship social network structure for simplicity of exposition. In order to develop our notation, the starting point is the linear model that assumes uncorrelated errors.

2.1 The Standard Model

The classical linear model for a population of N individuals can be written in matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \tag{1}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_N)'$ is a population vector of responses, $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_N)'$ with $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})'$ is the model design matrix for the population with p columns defined by a set of covariates that depend on auxiliary population information, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_N)'$ is the vector of population model residuals with $\epsilon_i \sim N(0, \sigma^2)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is the vector of regression coefficients. The population mean vector and population covariance matrix of $\mathbf{Y}$ are then $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{V} = \sigma^2\mathbf{I}_N$. Here $\mathbf{I}_N$ denotes the identity matrix of order N.

It is assumed that the matrix $\mathbf{X}$ defined by the auxiliary population information does not include variables related to social networks, so (1) does not use social network information.

2.2 The Mixed Model

Survey populations are often hierarchical, and can be characterised as grouped into clusters, with each cluster j accounted for by a cluster-specific random effect u_j in the model. A simple mixed model, i.e. a model characterised by fixed and random effects, for such data is

$$Y_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + u_j + \epsilon_{ij}; j = 1, \ldots, K; i = 1, \ldots, N_j \tag{2}$$

with $u_j \sim N(0, \sigma_a^2)$ and $\epsilon_i \sim N(0, \sigma_e^2)$. Here Y_{ij} is the value of the response for subject i (level 1) in cluster j (level 2), and we note that this model implies the following covariance structure

$$\text{Cov}(y_{ij}, y_{kl}) = \begin{cases} \sigma_u^2 + \sigma_e^2 & i = k \text{ and } j = l; \\ \sigma_u^2 & i \neq k \text{ and } j = l; \\ 0 & j \neq l. \end{cases}$$

It is well known that this mixed model can be written as the general linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim (\mathbf{0}, \boldsymbol{\Sigma}), \tag{3}$$

where $\boldsymbol{\Sigma} = \text{Diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K)$ has a block diagonal structure and corresponds to the covariance matrix with elements of $\boldsymbol{\Sigma}_i$ determined by $\text{Cov}(y_{ij}, y_{il})$.

As in Section 1.1 it is assumed that (3) does not use social network information. In the rest of this section we therefore describe linear models for the response variable that use a social network as an additional source of auxiliary information.

2.3 The Contextual Network (CN) Model

Consider an educational modelling exercise where academic performance (AP) is the response variable and socio-economic status (SES) of the student is the explanatory variable. A classical contextual approach might then lead one to include the average SES of the student's school as another explanatory variable. Friedkin (1990) adapts this idea to network data by considering models where the response for a particular subject also depends on the characteristics of other subjects that are linked to the one of interest. In our example this would correspond to modelling AP in terms of both the student's SES as well as the SES values of the student's friends. Since a student will generally have several friends, a student's AP could then be modelled in terms of his/her SES as well as the average SES of the his/her friends.

In general, such a CN model can be written in matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{4}$$

where $\mathbf{Y}$ and $\mathbf{X}$ have the same meaning as for model (1), but the columns of $\mathbf{U}$ correspond to statistics derived from the variables that are measured on the network. In particular, the ith row of $\mathbf{U}$ contains appropriate summary characteristics of those other individuals on the network that are linked to individual i. Thus, in the preceding example, assuming that SES is the only covariate

measured on the network, then $\mathbf{U}$ is the column vector of length N whose ith value is $\overline{SES}_i$, the average SES of all friends of student i. More generally, letting $\mathbf{T}$ denote the population matrix of covariates measured on the network, then one way of defining $\mathbf{U}$ is via the identity

$$\mathbf{U} = \mathbf{WT} \tag{5}$$

where $\mathbf{W}$ is a row-normalised version of $\mathbf{Z}$, i.e. the rows of $\mathbf{W}$ sum to one.

Remark

A contextual variable for person i often includes the value for this person, for example a household contextual effect is computed over all household members including person i. However, the contextual value for person i defined by (5) excludes person i, because $Z_{ii} = 0$ by definition.

2.4 The Autocorrelation (AR) Model

The matrix $\mathbf{T}$ introduced in the preceding description of the CN model can be any set of measurements on the individuals in the network. In particular, it can be $\mathbf{Y}$. This leads to another class of models, called Autocorrelation (AR) models, and also known as network effects models, that incorporate network information into a linear structure. See, for example, Doreian et al. (1984), Duke (1993), Marsden and Friedkin (1993) and Leenders (2002), and in the context of spatial models, Ord (1975). Under an AR model,

$$\mathbf{Y} = \theta\bar{\mathbf{Y}} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{6}$$

where $\bar{\mathbf{Y}} = (\bar{Y}_1, \ldots, \bar{Y}_N)'$ and $\bar{Y}_i$ is the average response of the individuals in the network that are linked to individual i, so $\bar{\mathbf{Y}} = \mathbf{WY}$, with $\mathbf{W}$ defined in the previous sub-section. The conditional (on $\mathbf{X}$) mean and variance of $\mathbf{Y}$ under (6) are $\boldsymbol{\mu} = \mathbf{D}^{-1}\mathbf{X}\boldsymbol{\beta}$ and $\mathbf{V} = \sigma^2(\mathbf{D}'\mathbf{D})^{-1}$, where $\mathbf{D} = \mathbf{I}_N - \theta\mathbf{W}$. Note that $\mathbf{W}$ can be defined in a variety of ways, see Leenders (2002), though typically it is defined as the row-normalised version of $\mathbf{Z}$, i.e. $\sum_{j=1}^{N} W_{ij} = 1$. The parameter θ is restricted

to the open interval $(-1, +1)$ as a necessary condition for $\mathbf{V}$ to exist.

In the context of the academic performance example introduced in the previous sub-section we see that (6) implies that a student's AP score now depends on his/her SES value as well as the average AP scores of his/her friends.

2.5 The Network Disturbance (ND) Model

Models of this type have been considered by Ord (1975) and Leenders (2002) among others, and correspond to imposing an AR structure on the error term in the standard linear model (1). They are referred to as network disturbance (ND) models and are specified by

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} = \theta\bar{\boldsymbol{\epsilon}} + \mathbf{v}, \text{where } \mathbf{v} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_N). \tag{7}$$

Here $\bar{\boldsymbol{\epsilon}} = (\bar{\epsilon}_1, \ldots, \bar{\epsilon}_N)$ where $\bar{\epsilon}_i$ is the average error of those individuals in the network linked to individual i. Returning to the academic performance example introduced in sub-section 2.3, the model can be interpreted as implying that if a student's friends have a below/above average AP value (as predicted by their SES values), then the student is more likely to have an AP value that is also below/above average.

Note that the model (7) can be re-written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2(\mathbf{D}'\mathbf{D})^{-1}) \tag{8}$$

where $\mathbf{D}$ was defined at the end of the previous sub-section, with $|\theta| < 1$. The parameter θ is an indicator of the strength of the between individual correlations generated by the network. For $\theta = 0$, the correlation between the Y values of any two individuals in the network is zero after one adjusts for their respective values of X. Under (8), the conditional (on $\mathbf{X}$) mean and variance of $\mathbf{Y}$ are $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{V} = \sigma^2(\mathbf{D}'\mathbf{D})^{-1}$ respectively.

It is worth pointing out that under the ND model, $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ is unaffected by the social network, whereas under the AR model (6), $\boldsymbol{\mu} = \mathbf{D}^{-1}\mathbf{X}\boldsymbol{\beta}$ depends on the network through $\mathbf{D}$. That is, under

the ND model, the expected value of Y for an individual only depends on the values of that individual's covariates. Unbiased prediction of Y can therefore ignore the network. Of course efficient prediction depends on the second order moments of (8), and so requires network information - as does prediction variance and mean squared error estimation. This is analogous to estimation under a multi-level model, where one can ignore the multi-level structure of the data if unbiased estimation is the aim, but one needs to take this structure into account for efficient inference.

2.6 Combining Network Effects and Area Effects

Under the assumption that residual heterogeneity can be modelled using hierachical random effects, the network models can be combined with area random effects. To start the mixed model (2) can also be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{G}\mathbf{u} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim (\mathbf{0}, \sigma_e^2\mathbf{I}), \tag{9}$$

with random intercept design matrix $\mathbf{G} = \text{Diag}(\mathbf{1}_{N_1}, \ldots, \mathbf{1}_{N_K})$, $\mathbf{u} = (u_1, \ldots, u_K)' \sim N(\mathbf{0}, \sigma_a^2\mathbf{I}_K)$, where $\mathbf{1}_n$ is a vector of ones of length n. Define the intra-cluster correlation by $\rho = \sigma_a^2/\sigma^2$ with $\sigma^2 = \sigma_a^2 + \sigma_e^2$. Then $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{V} = \sigma^2(\rho\mathbf{G}\mathbf{G}' + (1-\rho)\mathbf{I}_N)$.

Using this notation, the contextual model with area random effects can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\gamma} + \mathbf{G}\mathbf{u} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_N) \tag{10}$$

leading to $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\gamma}$ and $\mathbf{V} = \sigma^2(\rho\mathbf{G}\mathbf{G}' + (1-\rho)\mathbf{I}_N)$.

Similarly the AR and ND models with area random effects can be expressed as

$$\mathbf{Y} = \theta\bar{\mathbf{Y}} + \mathbf{X}\boldsymbol{\beta} + \mathbf{G}\mathbf{u} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_N) \tag{11}$$

leading to $\boldsymbol{\mu} = \mathbf{D}^{-1}\mathbf{X}\boldsymbol{\beta}$ and $\mathbf{V} = \sigma^2\mathbf{D}^{-1}(\rho\mathbf{G}\mathbf{G}' + (1-\rho)\mathbf{I}_N)(\mathbf{D}^{-1})'$, and

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{G}\mathbf{u} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2(\mathbf{D}'\mathbf{D})^{-1}), \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_N) \tag{12}$$

leading to $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{V} = \sigma^2\mathbf{D}^{-1}(\rho\mathbf{G}\mathbf{G}' + (1-\rho)\mathbf{I}_N)(\mathbf{D}^{-1})'$.

3 Prediction of Population Totals Using Network Models

The models discussed in the previous section are predictive models, i.e. when second order moments are known, they can be used to compute efficient predictions of unknown values of the response variable. We now describe how these models can be fitted, and how predicted values derived from them can be used to estimate the population total $T_Y = \sum_{i\in U} Y_i$ given the sample values $\{Y_i \quad i \in s\}$, the population matrix of model covariates $\mathbf{X}$ and either part of or all of the network matrix $\mathbf{Z}$. Throughout we assume that inclusion in sample does not depend on $\mathbf{Z}$ and that there is non-informative sampling given $\mathbf{X}$, see Section 1.4 in Chambers and Clark (2012). Consequently, all unknown parameter values for the standard model (1) can be estimated from the sample data and predicted values of Y for the non-sampled population individuals can be computed. We start by summarising known results from finite population estimation theory.

3.1 The Empirical Best Linear Unbiased Predictor

Let $\mathbb{E}\mathbf{Y} = \boldsymbol{\mu} = \mathbf{H}\boldsymbol{\lambda}$, where $\mathbf{H}$ is a known matrix with N rows and q columns and $\boldsymbol{\lambda}$ is an unknown parameter vector of length q. Also, suppose that $\text{Var}(\mathbf{Y}) = \mathbf{V}$ is a positive definite matrix of order N whose value is known up to a constant of proportionality. Examples of $\mathbf{H}$ and $\mathbf{V}$ are given in the following sub-section. The best linear unbiased predictor or BLUP of the population total $T_Y = \sum_{i\in U} Y_i$ is then an efficient estimator of this quantity, see Royall (1976). In order to specify the BLUP, let s and r denote the n sampled and $N-n$ non-sampled population individuals respectively, and put $\mathbf{H} = (\mathbf{H}_s', \mathbf{H}_r')'$ and $\mathbf{Y} = (\mathbf{Y}_s', \mathbf{Y}_r')'$. The matrix $\mathbf{V}$ can then be partitioned conformably as

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{ss} & \mathbf{V}_{sr} \\ \mathbf{V}_{rs} & \mathbf{V}_{rr} \end{pmatrix}.$$

A standard expression for the BLUP is its so-called predictive form

$$\hat{T}_Y^{BLUP} = \sum_{i \in s} Y_i + \sum_{i \in r} \mathbf{H}_i \hat{\boldsymbol{\lambda}} + \sum_{i \in s} \tau_i (Y_i - \mathbf{H}_i \hat{\boldsymbol{\lambda}}) \tag{13}$$

where $\mathbf{H}_i$ is the ith row of $\mathbf{H}$, $\hat{\boldsymbol{\lambda}} = (\mathbf{H}_s' \mathbf{V}_{ss}^{-1} \mathbf{H}_s)^{-1} \mathbf{H}_s' \mathbf{V}_{ss}^{-1} \mathbf{Y}_s$ is the best linear unbiased estimator (BLUE) of $\boldsymbol{\lambda}$, and τ_i is the ith element of the vector $\mathbf{V}_{ss}^{-1} \mathbf{V}_{sr} \mathbf{1}_{N-n}$, with $\mathbf{1}_{N-n}$ denoting a vector of ones of size $N - n$.

However, the BLUP can also be expressed as a weighted sum $\hat{T}_Y^{BLUP} = \sum_{i \in s} w_i Y_i = \mathbf{w}_s' \mathbf{Y}_s$ of the sample values of Y, where

$$\mathbf{w}_s = \mathbf{1}_n + \mathbf{V}_{ss}^{-1} \mathbf{A} \mathbf{B} \mathbf{1}_{N-n} \tag{14}$$

is the vector of BLUP weights. Here $\mathbf{1}_n$ is a vector of ones of size n, $\mathbf{A} = \mathbf{V}_{sr} \mathbf{1}_{N-n} - \mathbf{H}_s (\mathbf{H}_s' \mathbf{V}_{ss}^{-1} \mathbf{H}_s)^{-1}$ and $\mathbf{B} = \mathbf{H}_s' \mathbf{V}_{ss}^{-1} \mathbf{V}_{sr} - \mathbf{H}_r'$.

A key assumption of the BLUP is that the variance matrix $\mathbf{V}$ is known up to a constant of proportionality. This is often unrealistic, since $\mathbf{V}$ can depend on unknown parameters, which must then be estimated. Methods for doing this are described in the next section. Substituting these estimates into $\mathbf{V}$ defines its plug-in estimator $\hat{\mathbf{V}}$, which can be used in (14) instead of $\mathbf{V}$. The resulting estimator of the population total is called the empirical BLUP or EBLUP.

3.2 Calculating the EBLUP under Network Models

In order to use the EBLUP with the different network models defined in the previous section, we need to specify $\mathbf{H}$ and $\mathbf{V}$ as well as estimators of the unknown parameters that underpin these matrices. These are defined as follows:

Standard Model : Here $\mathbf{H} = \mathbf{X}$ and $\mathbf{V} = \sigma^2 \mathbf{I}_N$. The residual mean squared error defines an unbiased estimator of σ^2.

Mixed Model : Here $\mathbf{H} = \mathbf{X}$ and $\mathbf{V} = \boldsymbol{\Sigma}$. To obtain unbiased estimates of σ_a^2 and σ_e^2 that

define $\boldsymbol{\Sigma}$ restricted restricted maximum likelihood estimation (REML) can be applied using `R` package `lme4` (Bates et al. 2014).

CN Model : For this model $\mathbf{H} = [\mathbf{X}, \mathbf{U}]$ and $\mathbf{V} = \sigma^2\mathbf{I}_N$. We can unbiasedly estimate σ^2 using the residual mean squared error.

AR Model : In this case $\mathbf{H} = \mathbf{D}^{-1}\mathbf{X}$ with $\mathbf{D} = \mathbf{I}_N - \theta\mathbf{W}$ and $\mathbf{V} = \sigma^2(\mathbf{D}'\mathbf{D})^{-1}$. Estimates of σ^2 and θ can be obtained by maximum likelihood (ML). Restricted ML (REML) is often used to obtain unbiased variance estimates but it cannot be applied here, because both the mean and variance depend on the parameter θ. The EBLUP uses the plug-in estimates of $\mathbf{H}$ and $\mathbf{V}$ defined by the ML estimates of σ^2 and θ.

ND Model : Here $\mathbf{H} = \mathbf{X}$ and $\mathbf{V} = \sigma^2(\mathbf{D}'\mathbf{D})^{-1}$. ML estimation of σ^2 and θ can be carried out, and the resulting plug-in estimate of $\mathbf{V}$ used to calculate the EBLUP.

CN Model with area effects : $\mathbf{H} = [\mathbf{X}, \mathbf{U}]$ and $\mathbf{V} = \sigma^2(\rho\mathbf{G}\mathbf{G}' + (1-\rho)\mathbf{I}_N)$. REML estimation can be used as for the 'Mixed model' above.

AR Model with area effects : $\mathbf{H} = \mathbf{D}^{-1}\mathbf{X}$ and $\mathbf{V} = \sigma^2\mathbf{D}^{-1}(\rho\mathbf{G}\mathbf{G}' + (1-\rho)\mathbf{I}_N)(\mathbf{D}^{-1})'$. We have implemented ML estimation for this model.

ND Model with area effects : $\mathbf{H} = \mathbf{X}$ and $\mathbf{V} = \sigma^2\mathbf{D}^{-1}(\rho\mathbf{G}\mathbf{G}' + (1-\rho)\mathbf{I}_N)(\mathbf{D}^{-1})'$. We have implemented REML and ML estimation for this model.

ML estimation of σ^2 and θ for the AR and ND models is not straightforward. Both models are not reproducible, i.e. they do not share the property that the model for a subset of units of the population has the same form as the model for the whole population. To see this, note that the variance of the population response vector $\mathbf{Y}$ under both models is $\sigma^2(\mathbf{D}'\mathbf{D})^{-1}$ so that the variance for the sample response vector $\mathbf{Y}_s$ is $\sigma^2[(\mathbf{D}'\mathbf{D})^{-1}]_{ss}$. In general, this will not equal $\sigma^2(\mathbf{D}'_{ss}\mathbf{D}_{ss})^{-1}$, which is the assumed variance if the model is fitted via ML at the sample level. This misspecification can lead to biased estimates of the model parameters. A modified approach that yields

unbiased estimates of the fixed effects in the model is described in Suesse (2012). However this is computationally intensive. An alternative approach replaces $\mathbf{D}^{-1}$ by a 4th order Taylor series approximation. This speeds up computation considerably since it effectively replaces matrices of dimension $N \times N$ by matrices of dimension $n \times n$. See Suesse (2012) where it is shown that ML estimates based on this approximation are essentially identical to those obtained using the modified ML method.

3.3 Variance Estimation for the EBLUP

The prediction variance of the BLUP is

$$\text{Var}(\hat{T}^{BLUP} - T) = \tilde{\mathbf{w}}'\mathbf{V}\tilde{\mathbf{w}} \tag{15}$$

with $\tilde{\mathbf{w}} = (\mathbf{w}_s - \mathbf{1}_n, -\mathbf{1}_{N-n})$. This formula assumes that the vector of survey weights $\mathbf{w}_s$ is fixed. We can use the same formula for the EBLUP, although from (14) it is clear that the EBLUP weights are not fixed in general because the plug-in estimates of $\mathbf{H}$ and $\mathbf{V}$ used to calculate them will depend on estimated parameters. However, the increase in the prediction variance due to ML estimation of these parameters will be small for large sample sizes, and can be ignored, see Chambers et al. (2011).

Using (15) to estimate the prediction variance of the EBLUP depends on correct specification of the second order moments of Y. For the standard model and the CN model, we can avoid this by using an alternative prediction variance estimator that does not rely on specification of these second order moments, see Section 9.2 of Chambers and Clark (2012). This estimator is given by

$$\widehat{\text{Var}}(\hat{t}_{BLUP} - t) = \sum_{i \in s}(w_{is} - 1)^2(Y_i - \hat{\mu}_i)^2 + (N - n)\hat{\sigma}^2 \tag{16}$$

where $\hat{\mu}_i$ is the estimated mean for $i \in s$, i.e. $\hat{\mu}_i = \mathbf{X}_i\hat{\boldsymbol{\beta}}$ for the standard model and $\hat{\mu}_i = \mathbf{X}_i\hat{\boldsymbol{\beta}}+\mathbf{U}_i\hat{\boldsymbol{\gamma}}$ for the CN model, with $\hat{\sigma}^2$ corresponding to the usual unbiased estimator of σ^2 under each model.

For the AR and ND models we use equation (15) with a plug-in estimator $\hat{\mathbf{V}}$. In this context, we note that ML estimates of variance parameters are known to be biased, which could therefore lead to a bias in $\hat{\mathbf{V}}$ and in the resulting plug-in estimator defined by (15). The standard approach to dealing with this issue is to apply REML instead of ML. Unfortunately, the AR model does not allow the application of REML, and furthermore REML is computationally more complex when fitting these population models. Consequently a bias-corrected version of ML was applied, based on the approach set out in Goldstein (1989), which adjusts IGLS to obtain estimates that are equivalent to REML. The details of this are outlined in the Appendix of Suesse and Chambers (2014).

4 Modelling and Imputation of Networks

Our EBLUP development in the previous section assumed that the matrix $\mathbf{Z}$ defining the network is known. This is rather unlikely to be the case. It is far more likely that we will know either just that part of the network defined by the sampled individuals (i.e. $\mathbf{Z}_{ss}$) or that part of the network defined by the sampled individuals and their corresponding network links (i.e. $\mathbf{Z}_{ss}$ and $\mathbf{Z}_{sr}$). An implementation of a 'network-based' EBLUP in this situation must therefore take account of this incomplete network data. In this section we describe simple model-based imputation methods that can be used to approximate the impact of the unknown full network (i.e. $\mathbf{Z}$) on this EBLUP. In turn, this requires that we have a way of modelling $\mathbf{Z}$, given that we see only a part of this matrix. We start with a brief overview of models for networks.

4.1 Exponential Random Graph Models

The most popular class of models for a network $\mathbf{Z}$ is the class of (curved) exponential random graph models (ERGMs), these are discussed in Wasserman and Faust (1994) and Carrington et al.

(2005). Under an ERGM, the distribution of $\mathbf{Z}$ is characterised by

$$\Pr(\mathbf{Z} = \mathbf{z}) = \exp\left(\boldsymbol{\eta}(\boldsymbol{\zeta})'\mathbf{G}(\mathbf{z}) - \kappa(\boldsymbol{\zeta})\right), \tag{17}$$

where $\boldsymbol{\zeta}$ is the vector of model parameters, $\boldsymbol{\eta}(\boldsymbol{\zeta})$ is a mapping from p-dimensional to q-dimensional space with $p \leq q$, and $\kappa(\boldsymbol{\zeta})$ is the normalising constant. Here $\mathbf{G}(\mathbf{z})$ is a vector of q 'network statistics' which, together with $\boldsymbol{\zeta}$, completely characterises the distribution of $\mathbf{Z}$. Simple examples of network statistics are the number of 'edges' in the network (i.e. the number of observed links, usually expressed as a fraction of the total number $N(N-1)$ of possible links) and the number of triangles (a triangle is said to exist between individuals i, j and k, if $Z_{ij} = Z_{jk} = Z_{ik} = 1$). A more complicated, but widely used network statistic is GWESP, or the geometrically weighted edgewise shared partner statistic. Roughly speaking, this corresponds to a weighted sum, over possible values of m, of counts of the number of links 'connecting' any two individuals in the network who are themselves linked to exactly m other individuals. Like interaction terms in regression, such statistics allow one to model networks whose 'connectivety' structure is extremely complicated.

Fitting an ERGM via ML is usually not possible, mainly because direct calculation of the normalising constant $\kappa(\boldsymbol{\zeta})$ is infeasible. One way of circumventing this problem is to sample from the network distribution (17) using a Markov-Chain-Monte-Carlo (MCMC) algorithm in order to obtain a stochastic approximation to the maximum likelihood estimate of $\boldsymbol{\zeta}$. Such estimates are called MCMC ML estimates (Hunter and Handcock 2006). Describing the network distribution via simple network statistics, such as the number of triangles then becomes problematic, because such specifications often lead to degenerate MCMC samples. Some authors (Snijders 2002; Snijders et al. 2006) have therefore proposed the use of more complex network statistics, such as the family of GWESP statistics, for which degeneracy seems less of a problem. For more details of network modelling, see Strauss and Ikeda (1990); Hunter and Handcock (2006); Hunter (2007); Hunter et al. (2008b) and Butts (2008).

4.2 Imputation of Partly Observed Networks

An estimate $\hat{\mathbf{Z}}$ of the full network is necessary for calculation of the EBLUP under the network models considered in this paper. However, in practice only part of the network will be observed, say $\mathbf{Z}^{obs}$, and another part will be missing, say $\mathbf{Z}^{mis}$. For example, the observed network $\mathbf{Z}^{obs}$ could be $\mathbf{Z}_{ss}$, in which case the missing network $\mathbf{Z}^{mis}$ is $\mathbf{Z}_{sr} \cup \mathbf{Z}_{rs} \cup \mathbf{Z}_{rr}$. In what follows we assume an undirected network, i.e. $\mathbf{Z} = \mathbf{Z}'$, so $\mathbf{Z}_{sr} = \mathbf{Z}_{rs}$. We also focus on single-value imputation of $\mathbf{Z}^{mis}$. Our approach can be extended to multiple imputation.

Method 1

An efficient model-based approach to imputing the missing network components is to use the minimum mean square error predictor (MMSEP) $E(\mathbf{Z}^{mis}|\mathbf{Z}^{obs} = \mathbf{z}^{obs})$ under an appropriate ERGM for $\mathbf{Z}$. Unfortunately, these conditional expectations are often intractable and can only be estimated by sampling from the ERGM using MCMC methods. This approach is impractical in a simulation study and so we describe a simpler, more feasible, approach.

Suppose conditionally on $\mathbf{z}^{obs}$ that Z_{ij}^{mis} and Z_{kl}^{mis} are conditionally independent for any two distinct pairs of individuals i, j and k, l, where both pairs are in mis and by distinct we mean that $(i, j) \neq (k, l)$ and $(i, j) \neq (l, k)$ hold. Then $\Pr(\mathbf{Z}^{mis} = \mathbf{z}^{mis}|\mathbf{Z}^{obs} = \mathbf{z}^{obs}) = \prod_{ij \in mis} \Pr(Z_{ij} = z_{ij}|\mathbf{Z}^{obs} = \mathbf{z}^{obs})$, where $\prod_{ij \in mis}$ denotes the product over all distinct pairs is in the set mis. It follows that we can write, for a distinct pair $(i, j) \in mis$,

$$\frac{\Pr(Z_{ij} = 1|\mathbf{Z}^{obs} = \mathbf{z}^{obs})}{\Pr(Z_{ij} = 0|\mathbf{Z}^{obs} = \mathbf{z}^{obs})} = \exp(\boldsymbol{\eta}^T(\boldsymbol{\zeta})\Delta\mathbf{G}_{ij}^{mis}), \tag{18}$$

where $\Delta\mathbf{G}_{ij}^{mis}$ is the change statistic, i.e. the difference in $\mathbf{G}$ between

$$(Z_{ij}, \mathbf{z}^{mis-(i,j)}, \mathbf{z}^{obs}) = (1, \mathbf{z}^{mis-(i,j)}, \mathbf{z}^{obs})$$

and

$$(Z_{ij}, \mathbf{z}^{mis-(i,j)}, \mathbf{z}^{obs}) = (0, \mathbf{z}^{mis-(i,j)}, \mathbf{z}^{obs}).$$

Note that $mis - (i,j)$ here denotes the set mis with the distinct pair (i,j) excluded. Re-arranging Equation (18) gives the MMSEP under conditional independence, $E(Z_{ij} = 1|\mathbf{Z}^{obs} = \mathbf{z}^{obs}) = \Pr(Z_{ij} = 1|\mathbf{Z}^{obs} = \mathbf{z}^{obs}) = \text{expit}(\boldsymbol{\eta}^T(\boldsymbol{\zeta})\Delta\mathbf{G}_{ij}^{mis})$ with $\text{expit}(x) = \exp(x)/(1+\exp(x))$. It remains to observe that it is only necessary to compute $\Delta\mathbf{G}_{ij}^{mis}$ in order to obtain this MMSEP for any distinct pair $(i,j) \in mis$. Since the conditional independence assumption is generally unwarranted, this approach can only be considered as defining an approximation to $\Pr(\mathbf{Z}^{mis}|\mathbf{Z}^{obs} = \mathbf{z}^{obs})$. However, it is computationally feasible for realistic sample and population sizes. This imputation method equires knwoledge of the ERGM parameters. The `R` package `ergm` (Hunter et al. 2008a) can fit ERGMs with missing data. However fitting ERGMs with a lage portion of missing data, takes much longer (on a single core of a modern High Performance Cluster 4 hours) than fitting a model to a network with fully observed data (a few seconds). This means fitting an ERGM in a simulation study for thousands of data sets is not feasible. Instead we use the true known $\boldsymbol{\zeta}$.

Method 2

An even simpler approach is to calculate the proportion of $Z_{ij} = 1$ in $\mathbf{z}^{obs}$ and use this proportion (the network density) to impute $\mathbf{Z}^{mis}$. This corresponds to imputing on the basis of an ERGM model defined by just the EDGES statistic, i.e. the number of edges in the network. This ERGM is equivalent to assuming that each Z_{ij} in the network matrix $\mathbf{Z}$ is an independent Bernoulli variable with a common probability of a 'success'.

5 Simulation Study

5.1 Study Design

This section contains the results from a simulation study whose aim was to investigate the effect of using networks as an additional source of information when estimating the population total T_Y of a survey variable Y. A networked population of size $N = 1,000$ was independently simulated 2,000 times, balancing computation time against the number of different scenarios that were explored in the study, and independent simple random samples of size $n = 100$ and $n = 200$ were independently selected without replacement from each simulated population.

Network Generation

The literature on network analysis suggests that networks are often well characterised by an ERGM defined in terms of an EDGES (number of edges) statistic and a GWESP statistic (Hunter et al. 2008a). Consequently, $\mathbf{Z}$ was generated as a random draw from an ERGM with an EDGES statistic equal to θ on the logit scale and a weight parameter of 1.0 for the GWESP statistic. In what follows we use ERGM(m) to denote such an ERGM, where m is the network density, i.e. the average number of links per individual. The values $\theta = -5.81, -4.18, -2.944$ were then chosen in order to generate a network with a density of about $m = 3, 15, 50$ network links respectively for each individual, i.e. $m \approx P(Z_{ij} = 1) \times N$ with $P(Z_{ij} = 1) \approx \text{expit}(\theta)$. Note that with this specification the number of network links for an individual is random, with only the approximate population average number of links fixed. However we also consider the situation where we constrain the number of links to 3 links per individual following the 'three best friends' BHPS example. This is achieved by removing or adding links randomly to each individual until the desired 3 links are obtained.

Imputation of Partly Observed Networks

In the simulation study we restricted ourselves to the two most realistic cases where network data are observed on a sample. In both cases, only part of the network is observed and so the unobserved components must be imputed. In the first case, denoted by SS in what follows, only $\mathbf{Z}_{ss}$ is observed and so $\mathbf{Z}_{sr}$ and $\mathbf{Z}_{rr}$ are missing. In the second case, denoted by SS+SR in what follows, $\mathbf{Z}_{ss}$ and $\mathbf{Z}_{sr}$ are observed but $\mathbf{Z}_{rr}$ is missing. This second case is more realistic from the viewpoint of having usable network information, since here we at least have complete network information for all sampled individuals.

For the ERGM network both imputation Method 1 and imputation Method 2 lead to the same imputed value of $\mathbf{Z}^{mis}$ in the SS scenario. In contrast, these methods lead to different imputed values in the SS+SR scenario under the ERGM network. We therefore denote the application of imputation Method 1 in the SS+SR scenario for the ERGM network by SS+SR/1, and the corresponding application of imputation Method 2 by SS+SR/2.

Finally, we also considered the situation where no network data are used (the standard model and mixed model) and also the case where the population network is fully known.

Parameter Specification for Linear Network Models

We generated data under all four of the linear network models discussed in Section 2 with ($\rho = 0.1$) and without area random effects ($\rho = 0.0$).

In each case, all of these models were fitted to the sample data, and EBLUP estimates of the population total T were then computed based on these fits (see the discussion in Section 3). Even though we simulated population data under all the models described in Section 2.6, i.e. including both area and network effects, we did not fit network models with area effects to the sample data. This was done in order to simplify the simulation study and also because as will be seen from the results, accounting for the random effects does not yield significant efficiency gains.

Population data were simulated assuming $\sigma^2 = 2$, $\beta_0 = 40$ and $\beta_1 = 5$. Furthermore, two models for the auxiliary variable were considered: i) X takes values randomly in the set $X_i =$

$1, \ldots, 9$, and ii) $X_i \sim N(0, 25)$. Both models have high predictive power, since for case i) the Standard model implies a theoretical value of $R^2 = 0.928$, while for case ii) $R^2 = 125/127 = 0.984$ under this model.

CN Model :

$$Y_i = \beta_0 + X_i\beta_1 + U_i\gamma + \epsilon_i,\ \epsilon_i \sim N(0, \sigma^2)$$

Here $\gamma = 2$ and the contextual variable U_i is defined as the average value of X for all individuals in the network that are linked to individual i, i.e. $\mathbf{U} = \mathbf{WX}$, where $\mathbf{W}$ is the row-normalised version of $\mathbf{Z}$ and $\mathbf{X}$ denotes the vector of population values of X.

AR Model :

$$\mathbf{Y} = \theta\mathbf{WY} + \beta_0 + \mathbf{X}\beta_1 + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N(0, \sigma^2\mathbf{I}_N)$$

with $\theta = 0.5$.

ND Model :

$$\mathbf{Y} = \beta_0 + \mathbf{X}\beta_1 + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} = \theta\mathbf{W}\boldsymbol{\epsilon} + \mathbf{v},\ \mathbf{v} \sim N(0, \sigma^2\mathbf{I}_N)$$

with $\theta = 0.5$.

CN Model with area random effects : see Model (10) with $\rho = 0.1$, $\gamma = 2$.

AR Model with area random Effects : see Model (11) with $\rho = 0.1$ and $\theta = 0.5$.

ND Model with area random effects : see Model (12) with $\rho = 0.1$ and $\theta = 0.5$.

5.2 Simulation Results

Results for the $n = 100$ case with $\rho = 0.0$ and $\rho = 0.1$ and $X \sim U\{1, 2, \ldots, 9\}$ are presented. Tables 1, 2, and 3 show the Monte Carlo relative mean squared errors of the estimates of T when the network is generated under an ERGM where the total number of friends is random, with expectations 15, 3 and 50 respectively, and Table 4 shows the same relative mean squared errors when

this value is fixed at 3. Corresponding simulation results for $X \sim N(0, 25)$ are presented in Tables 5 to 8 in the Appendix. Results for the $n = 200$ case are similar. Note that we do not show Monte Carlo bias, since these values were effectively zero for all methods. The results displayed in each Table include the two cases where the network is ignored (the 'standard' model and the 'mixed' model) and when the population network matrix $\mathbf{Z}$ is fully known ('full network known'). For partially observed network data we show results for the SS case (only $\mathbf{Z}_{ss}$ known), the SS+SR/1 case ($\mathbf{Z}_{ss}$ and $\mathbf{Z}_{sr}$ known, Method 1 imputation) and the SS+SR/2 case ($\mathbf{Z}_{ss}$ and $\mathbf{Z}_{sr}$ known, Method 2 imputation). All results are shown relative to those for the BLUP, which uses complete network information as well as knowledge of θ. Although the level of knowledge required to compute the BLUP is unrealistic in practice, its performance provides us with a benchmark against which to gauge the relative benefit of putting more effort into collecting more network information and in carrying out more intensive network modelling for imputation of the unknown parts of the network. Furthermore, comparisons with the 'Standard' or 'Mixed' cases allow us to assess how much efficiency is lost by ignoring network information.

It is clear from the results shown in Tables 1 to 4 that ignoring the network (i.e. using the 'Standard' or 'Mixed' models for estimation) can lead to a large loss in efficiency if in fact either the AR or the CN models are true. Interestingly, our results also seem to indicate that adopting the CN model when in fact the AR model is true seems as good as using the correctly specified AR model when the number of friends is not small. Note that when the ND model is true, ignoring the network information in the data only leads to a marginal loss in efficiency. In fact, the EBLUPs based on the different network models are all almost fully efficient in this case, irrespective of whether the assumed network model is true. These results also indicate that there is very little difference between using the 'Mixed' or the 'Standard' models for estimation.

When $\mathbf{Z}$ is known, but not θ, we see a loss of efficiency of up to 97% under the AR model, mainly because the pseudo-design matrix $\mathbf{D}^{-1}(\theta)\mathbf{X}$ for this model depends on the estimated value of θ. As the number of friends increases, this loss of efficiency associated with having to estimate θ from the sample data decreases in importance. This problem is much less of an issue for the ND

model because in this case the design matrix does not depend on θ. Obviously, there is no impact under the CN model.

In order to see why the CN model yields similar results as the AR model when in fact the AR model holds, we note that the mean of the AR model is $\boldsymbol{\mu} = \mathbf{D}(\theta)^{-1}\mathbf{X}\boldsymbol{\beta}$. If we approximate $\mathbf{D}(\theta)^{-1}$ by a first order Taylor series around zero, i.e. $\mathbf{D}(\theta)^{-1} = (\mathbf{I}_N - \theta\mathbf{W})^{-1} \approx \mathbf{I}_N + \theta\mathbf{W}$, then

$$\boldsymbol{\mu} \approx \mathbf{X}\boldsymbol{\beta} + \theta\mathbf{W}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\gamma}$$

with $\boldsymbol{\gamma} = \theta\boldsymbol{\beta}$ and $\mathbf{U} = \mathbf{W}\mathbf{X}$. That is, the implied mean structure under the AR model is approximately the same as that under a CN model.

When $\mathbf{Z}_{ss}$ and $\mathbf{Z}_{sr}$ are observed, the EBLUP based on the CN model appears to perform well generally. This is because the EBLUP under this model does not depend on $\mathbf{Z}_{rr}$ and hence is unaffected by imputation of this part of the network. This is in contrast to the performance of this EBLUP when only $\mathbf{Z}_{ss}$ is observed. Here we see that the need to impute $\mathbf{Z}_{sr}$ leads to a significant loss of efficiency. Since estimation of θ in the pseudo-design matrix $\mathbf{D}(\theta)^{-1}\mathbf{X}$ under the AR model has a larger negative effect than the approximation of the AR model by the CN model, we conclude that the EBLUP based on the CN model seems a generally more robust method for estimating the population total than the EBLUP based on the AR model. The AR model's performance is good provided the number of friends is medium or large or the number of friends is fixed, however the model is not useful when the number of friends is small and not fixed, supporting the argument that the CN model is generally more robust and preferred, as it always provides efficiency gains in the SS+SR case.

Our results indicate that the imputation method SS+SR/1, based on the conditional independence method (Method 1), is never more efficient than SS+SR/2, the simple proportion approach (Method 2). While this result is somewhat surprising, it can be explained by the roughness of the approximation implicit in the imputes generated using SS+SR/1 and the robustness of the simple proportion method used in SS+SR/2. A priori, however, we would expect that a model-based ap-

proach using $\mathbb{E}(\mathbf{Z}^{mis}|\mathbf{Z}^{obs} = \mathbf{z}^{obs})$ as a predictor, say obtained by the MCMC technique, should lead to higher efficiency than using SS+SR/2. In particular, it is possible that a multiple imputation version of SS+SR/1 might yield better results. However such methods still need to be developed for ERGMs, see for example Koskinen et al. (2011). In any case, it should be noted that the small observed differences between SS+SR/2 and the complete network known case indicates that the possible gains from the use of these more sophisticated methods may be minimal.

We now focus on the case where the expected number of friends per subject is small, here equal to 3. When the actual number of friends is random, see Table 2, there is no gain associated with using imputation method SS compared to ignoring the network information and basing estimation on the 'Standard' or 'Mixed' models. Here we see that network imputation based on SS+SR/1 and SS+SR/2 provide some gains relative to ignoring the network when the contextual CN model is fitted, but still lead to a loss in efficiency when the AR model is fitted and the AR model holds, and only a small gain when the AR model is fitted but the CN model holds. Overall, it seems that when the number of friends is random, with a small expected number of friends, then it is not worthwhile to apply the AR model, whereas there is still value in fitting a CN model. Table 4 on the other hand shows what happens when the number of friends remains small but is now fixed. Here we see a dramatic improvement in performance for methods based on the CN and AR models, with results similar (but with greater relative mean squared errors) to the situations where the number of friends is random, but with a larger expected value. This 'fixing' eliminates variation in the calculation of the matrix $\mathbf{W}$ which requires dividing by the unknown number of friends. It also leads to more efficient imputation, explaining the superior performance of the CN and AR model when the number of friends is fixed compared to this number being random. For example when the number of friends is fixed at three and a sampled subject has 2 friends inside the sample, then we know that there is exactly 1 remaining friend in the non-sampled part of the population. The matrix $\mathbf{W}$ is then simply the imputed $\mathbf{Z}$ divided by 3.

Average lengths and associated coverages for nominal 95% Gaussian confidence intervals generated by the estimates of the mean squared errors of the different estimators are set out in Tables

Table 1: $n = 100$: ERGM(15) network & $X \sim U\{1, .., 9\}$: Ratio of MSE(EBLUP) to MSE(BLUP)

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP - actual MSE		$19,333$	$18,810$	$18,803$	$16,882$	$18,131$	$18,130$
$\mathbf{Z}_U$ known ($\rho = 0.0$)		1.00	1.00	1.00	1.02	1.05	1.03
CN ($\rho = 0.0$)	SS	1.69	2.14	1.02	1.82	2.12	1.04
	SS+SR/1	1.10	1.16	1.00	1.14	1.16	1.04
	SS+SR/2	1.06	1.11	1.01	1.10	1.09	1.05
AR ($\rho = 0.0$)	SS	1.71	2.16	1.02	1.84	2.13	1.03
	SS+SR/1	1.11	1.15	1.00	1.14	1.15	1.04
	SS+SR/2	1.06	1.10	1.01	1.10	1.08	1.05
ND ($\rho = 0.0$)	SS	1.77	2.26	1.00	1.88	2.24	1.03
	SS+SR/1	1.78	2.25	1.00	1.89	2.21	1.03
	SS+SR/2	1.78	2.25	1.00	1.89	2.21	1.03
Standard ($\rho = 0.0$)		1.74	2.24	1.00	1.88	2.23	1.02
Mixed ($\rho = \hat{\rho}$)		1.76	2.26	1.02	1.91	2.24	1.00

10, 12, 14 and 16 for $X \sim U\{1, 2, \ldots, 9\}$, and Tables 9, 11, 13 and 15 for $X \sim N(0, 25)$, which can all be found in the Appendix. Monte Carlo coverages in all cases are close to the nominal level. However, the average confidence interval length in the SS+SR case is considerably shorter than that for the SS case when estimation is carried out under the AR and CN models. This provides further support for the conclusion reached above, that basing an EBLUP on a CN model seems a generally robust approach to using network data when estimating a population total, even though one must keep in mind that the simple linearization-based prediction variance estimator (16) used with the CN model slightly underestimates random variation due to its assumption of fixed sample weights, resulting in too narrow confidence intervals with slight undercoverage.

Table 2: $n = 100$: ERGM(3) network & $X \sim U\{1, \ldots, 9\}$: Ratio of MSE(EBLUP) to MSE(BLUP)

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP - actual MSE		19,190	22,686	22,638	17,137	22,320	22,352
$\mathbf{Z}_U$ known ($\rho = 0.0$)		1.00	1.01	1.03	1.02	1.06	1.04
CN ($\rho = 0.0$)	SS	5.88	9.66	1.06	6.58	10.5	1.10
	SS+SR/1	1.42	2.29	1.04	1.52	2.53	1.10
	SS+SR/2	1.42	2.29	1.04	1.51	2.52	1.10
AR ($\rho = 0.0$)	SS	6.20	10.1	1.05	6.85	10.9	1.10
	SS+SR/1	1.44	2.16	1.05	1.57	2.33	1.10
	SS+SR/2	1.44	2.18	1.05	1.58	2.32	1.10
ND ($\rho = 0.0$)	SS	6.21	10.4	1.04	6.88	11.2	1.05
	SS+SR/1	6.10	8.68	1.01	6.68	9.30	1.04
	SS+SR/2	6.15	8.67	1.01	6.64	9.39	1.04
Standard ($\rho = 0.0$)		6.11	10.1	1.05	6.85	10.9	1.09
Mixed ($\rho = \hat{\rho}$)		6.11	10.2	1.06	6.87	10.9	1.07

Table 3: $n = 100$: ERGM(50) network & $X \sim U\{1, .., 9\}$: Ratio of MSE(EBLUP) to MSE(BLUP)

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP - actual MSE		19,067	17,728	17,730	16,949	17,154	17,152
$\mathbf{Z}_U$ known ($\rho = 0.0$)		1.00	1.00	1.00	1.03	1.02	1.02
CN ($\rho = 0.0$)	SS	1.22	1.31	1.01	1.25	1.32	1.04
	SS+SR/1	1.15	1.20	1.01	1.17	1.21	1.03
	SS+SR/2	1.02	1.06	1.01	1.04	1.07	1.04
AR ($\rho = 0.0$)	SS	1.22	1.29	1.01	1.24	1.29	1.03
	SS+SR/1	1.15	1.20	1.00	1.17	1.20	1.03
	SS+SR/2	1.03	1.05	1.01	1.04	1.05	1.04
ND ($\rho = 0.0$)	SS	1.23	1.31	1.00	1.25	1.34	1.02
	SS+SR/1	1.23	1.31	1.00	1.25	1.34	1.02
	SS+SR/2	1.23	1.31	1.00	1.25	1.34	1.02
Standard ($\rho = 0.0$)		1.22	1.31	1.00	1.25	1.33	1.02
Mixed ($\rho = \hat{\rho}$)		1.23	1.31	1.00	1.25	1.33	1.02

Table 4: $n = 100$: ERGM(3) network with fixed number of friends & $X \sim U\{1, .., 9\}$: Ratio of MSE(EBLUP) to MSE(BLUP)

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP - actual MSE		$19,328$	$21,962$	$21,938$	$17,059$	$21,534$	$21,411$
$\mathbf{Z}_U$ known ($\rho = 0.0$)		1.00	1.03	1.02	1.02	1.12	1.05
CN ($\rho = 0.0$)	SS	4.43	6.84	1.02	4.97	6.80	1.09
	SS+SR/1	1.31	2.07	1.02	1.40	2.11	1.09
	SS+SR/2	1.31	2.07	1.02	1.40	2.11	1.09
AR ($\rho = 0.0$)	SS	4.67	7.02	1.01	5.25	6.97	1.09
	SS+SR/1	1.34	2.00	1.02	1.44	1.99	1.09
	SS+SR/2	1.34	2.00	1.02	1.44	1.99	1.09
ND ($\rho = 0.0$)	SS	4.95	7.60	1.02	5.44	7.44	1.05
	SS+SR/1	5.10	7.38	1.03	5.50	7.18	1.05
	SS+SR/2	5.10	7.38	1.03	5.50	7.18	1.05
Standard ($\rho = 0.0$)		4.86	7.64	1.00	5.41	7.53	1.08
Mixed ($\rho = \hat{\rho}$)		4.85	7.66	1.01	5.38	7.62	1.07

6 Discussion

At the end of Section 1, we stated that our aim in this paper is to address the questions: (i) Is embedding network information useful for survey estimation? (ii) If the answer to (i) is yes, then which models are potentially useful? and (iii) How much network data needs to be collected in order to obtain potentially higher precision for survey estimation? Given the simulation results that we present in Section 5, our tentative answer to (i) is yes, and our corresponding answer to (ii) is the CN and AR models when either model is true, because in both cases the mean of the response depends on the network. Our simulation results provide some evidence that this conclusion holds for almost all situations except where the number of links is small and random, in which case they suggest that only the CN model provides efficiency gains. However when the mean does not depend on the network, as is the case under the ND model, our results suggest that ignoring the network does not result in a significant loss of efficiency. We have also investigated this for other 'network covariance' models, where the mean structure is unaffected by the network, and we have observed similar results, see Suesse and Chambers (2014). In effect, ignoring the network under

the CN and AR models leads to a mis-specification of the mean model, but this does not apply for the ND (and similar) models. Finally, our answer to (iii) is that in realistic applications it will usually be impossible to collect the full network, and our simulation results are some evidence that when either the CN model or the AR model is true then both $\mathbf{Z}_{ss}$ and $\mathbf{Z}_{sr}$ must be collected in order to obtain efficiency gains. Knowledge of $\mathbf{Z}_{ss}$ alone is not enough.

In practice, we suggest a careful model fitting exercise be carried out before attempting to use either the CN model or the AR model for survey estimation. Given the numerical difficulties with fitting the AR model, see Suesse (2012), we recommend that the CN model be used if it is a reasonable fit to the data, otherwise caution is warranted and ignoring the network might be the best option.

Clearly, more extensive information on networks needs to be collected in conjunction with standard survey data to gain further insight into the usefulness of network models for survey estimation. In this paper we have focused on undirected networks, so knowing $\mathbf{Z}_{sr}$ is equivalent to knowing $\mathbf{Z}_{rs}$. For directed networks, this equivalence does not apply and conclusions, particularly for the case when $\mathbf{Z}_{ss}$ and $\mathbf{Z}_{sr}$ are known, are likely to be different. We make a start on the issue of imputation methods for the missing network information in this paper, but many questions remain. Is an appropriate single value imputation (let alone multiple imputation) method using $\mathbb{E}(\mathbf{Z}^{mis}|\mathbf{Z}^{obs} = \mathbf{z}^{obs})$ (Method 1) better than the simple proportion approach (Method 2)? The numerical intensity of the MCMC methods used to fit network models like the ERGM when population sizes are large meant that we could not fully explore this issue. There is current research that tries to address some of these issues (Koskinen et al. 2011), but more is required. However, given that the simple SS+SR/2 imputation method that ignores the fitted ERGM model is consistently better than the SS+SR/1 imputation method that uses the fitted ERGM, we anticipate that more sophisticated imputation methods are unlikely to lead to substantial efficiency gains in most cases. However in some cases, see Table 2, substantial gains seem possible.

All network models considered in this paper assume that the value of the response variable Y for an individual in the study population depends on a linear combination of the values of this

variable for the other individuals in the population that are linked to this person in the network. If there is an implicit ordering in the strength of these links, then this can be allowed for in the network model for Y. For example, in the case of a 'best friend' network, where the friendships are ordered by their strength, one can modify the CN model so that there is a separate parameter for each level of 'best friend', see Friedkin (1990) for similar examples. To illustrate, in the BHPS application, when this extended contextual model is fitted, a Wald test for equality of these effects supports the assumption of a common effect.

Finally, we note that throughout this paper we have assumed that the method of sampling is independent of the network structure given the available population auxiliary information. In effect, we assume that measurement of the network is something that is done on the sample (as in our BHPS application), rather than sampling being something that is carried out on the network. However, there are important applications, see Thompson and Seber (1996), where inclusion in sample depends on being linked to another sampled individual via a network. It is clear that in these cases we cannot treat the observed network structure in $\mathbf{Z}_{ss}$ and $\mathbf{Z}_{sr}$ in the same way as we have in this paper, and this 'informative' method of sampling needs to be taken into account when we attempt to impute the unknown components of $\mathbf{Z}$. Work on this problem is continuing.

References

Bates, D., Maechler, M., B., B., and Walker, S. (2014), "Linear mixed-effects models using Eigen and S4," .

Butts, C. (2008), "network: A Package for Managing Relational Data in R," *Journal of Statistical Software*, 24, 1–36.

Carrington, P., Scott, J., and Wasserman, S. (2005), *Models and methods in social network analysis*, New York: Cambridge University Press.

Chambers, R., Chandra, H., and Tzavidis, N. (2011), "On bias-robust mean squared error estimation for pseudo-linear small area estimators," *Survey Methodology*, 37, 153–170.

Chambers, R. L. and Clark, R. G. (2012), *An Introduction to Model-Based Survey Sampling with Applications*, Oxford: Oxford University Press.

Doreian, P., Teuter, K., and Wang, C. H. (1984), "Network auto-correlation models - some monte-carlo results," *Sociological Methods & Research*, 13, 155–200.

Duke, J. B. (1993), "Estimation of the network effects model in a large data set," *Sociological Methods & Research*, 21, 465–481.

Frank, O. and Strauss, D. (1986), "Markov Graphs," *Journal of the American Statistical Association*, 81, 832–842.

Friedkin, N. E. (1990), "Social networks in structural equation models," *Social Psychology Quarterly*, 53, 316–328.

Goldstein, H. (1989), "Restricted unbiased iterative generalized least-squares estimation," *Biometrika*, 76, 622–623.

Hunter, D., Handcock, M., Butts, C., Goodreau, S., and Morris, M. (2008a), "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks," *Journal of Statistical Software*, 24, 1–29.

Hunter, D. R. (2007), "Curved exponential family models for social networks," *Social Networks*, 29, 216–230.

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008b), "Goodness of fit of social network models," *Journal of the American Statistical Association*, 103, 248–258.

Hunter, D. R. and Handcock, M. S. (2006), "Inference in Curved Exponential Family Models for Networks," *Journal of Computational and Graphical Statistics*, 15, 565–583.

Koskinen, J., Robins, G., and Pattison, P. (2011), "Missing data in social networks: Problems and prospects for model-based inference," Working Paper No. 09-01, MelNet Social Networks Laboratory.

Leenders, R. (2002), "Modeling social influence through network autocorrelation: constructing the weight matrix," *Social Networks*, 24, 21–47.

Marsden, P. V. and Friedkin, N. E. (1993), "Network studies of social-influence," *Sociological Methods & Research*, 22, 127–151.

Ord, K. (1975), "Estimation methods for models of spatial interaction," *Journal of the American Statistical Association*, 70, 120–126.

Royall, R. M. (1976), "Linear least-squares prediction approach to 2-stage sampling," *Journal of the American Statistical Association*, 71, 657–664.

Snijders, T. (2002), "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models," *Journal of Social Structure*, 1–40.

Snijders, T., Pattison, P., Robins, G., and Handcock, M. (2006), "New Specifications for Exponential Random Graph Models," *Sociological Methodology*, 36, 99–153.

Srndal, C., Swensson, B., and Wretman, J. (1992), *Model assisted survey sampling*, Springer series in statistics, New York: Springer-Verlag.

Strauss, D. and Ikeda, M. (1990), "Pseudolikelihood Eestimation for Social Networks," *Journal of the American Statistical Association*, 85, 204–212.

Suesse, T. (2012), "Estimation in Autoregressive Population Models," in *Proceedings of Fifth Annual ASEARC Research Conference*, ASEARC, 2-3 February 2012.

Suesse, T. and Chambers, R. (2014), "Using Social Network Information for Survey Estimation," Tech. rep., National Institute of Applied Statistics Research Australia, University of Wollongong, technical Report.

Thompson, S. and Seber, G. (1996), *Adaptive sampling*, Wiley Series in probabilitu and mathematical statistics, New York: Wiley.

Wasserman, S. and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, New York: Cambridge University Press.

A Further Simulation Results

Table 5: $n = 100$: ERGM(15) network & $X \sim N(0, 25)$: Ratio of MSE(EBLUP) to MSE(BLUP)

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP - actual MSE		$18,217$	$19,690$	$19,703$	$17,763$	$17,682$	$17,706$
$\mathbf{Z}_U$ known ($\rho = 0.0$)		1.00	1.01	1.02	1.05	1.05	1.06
CN ($\rho = 0.0$)	SS	4.33	6.19	1.03	4.60	6.81	1.07
	SS+SR/1	1.53	1.94	1.03	1.56	1.99	1.05
	SS+SR/2	1.28	1.61	1.03	1.32	1.68	1.06
AR ($\rho = 0.0$)	SS	4.38	6.25	1.03	4.68	6.85	1.07
	SS+SR/1	1.52	1.86	1.03	1.56	1.93	1.05
	SS+SR/2	1.29	1.57	1.03	1.33	1.64	1.06
ND ($\rho = 0.0$)	SS	4.69	6.64	1.02	4.93	7.31	1.05
	SS+SR/1	4.66	6.50	1.02	4.91	7.14	1.05
	SS+SR/2	4.65	6.50	1.02	4.90	7.15	1.05
Standard ($\rho = 0.0$)		4.65	6.61	1.02	4.88	7.30	1.05
Mixed ($\rho = \hat{\rho}$)		4.65	6.64	1.02	4.89	7.35	1.02

Table 6: $n = 100$: ERGM(3) network & $X \sim N(0, 25)$: Ratio of MSE(EBLUP) to MSE(BLUP)

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP - actual MSE		18,092	23,647	23,697	17,911	24,361	24,469
$\mathbf{Z}_U$ known ($\rho = 0.0$)		1.00	1.03	1.03	1.05	1.12	1.04
CN ($\rho = 0.0$)	SS	22.4	38.7	1.04	22.9	37.3	1.05
	SS+SR/1	3.18	7.39	1.04	3.23	6.78	1.06
	SS+SR/2	3.17	7.37	1.04	3.22	6.76	1.06
AR ($\rho = 0.0$)	SS	23.7	40.7	1.04	24.3	39.2	1.05
	SS+SR/1	3.23	6.89	1.04	3.44	6.28	1.06
	SS+SR/2	3.24	6.82	1.04	3.33	6.34	1.06
ND ($\rho = 0.0$)	SS	23.9	41.4	1.04	24.2	39.8	1.04
	SS+SR/1	22.3	34.6	1.02	22.9	33.4	1.04
	SS+SR/2	22.3	34.6	1.02	22.9	33.3	1.04
Standard ($\rho = 0.0$)		23.5	40.3	1.04	23.9	38.7	1.05
Mixed ($\rho = \hat{\rho}$)		23.8	40.9	1.05	24.4	39.4	1.04

Table 7: $n = 100$: ERGM(50) network & $X \sim N(0, 25)$: Ratio of MSE(EBLUP) to MSE(BLUP)

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP - actual MSE		18,169	17,667	17,674	17,951	19,358	19,360
$\mathbf{Z}_U$ known ($\rho = 0.0$)		1.00	1.02	1.01	1.05	1.03	1.02
CN ($\rho = 0.0$)	SS	1.91	2.42	1.02	2.01	2.38	1.03
	SS+SR/1	1.62	1.88	1.02	1.63	1.78	1.03
	SS+SR/2	1.06	1.15	1.03	1.14	1.13	1.03
AR ($\rho = 0.0$)	SS	1.91	2.43	1.02	2.02	2.38	1.03
	SS+SR/1	1.60	1.85	1.02	1.62	1.76	1.04
	SS+SR/2	1.06	1.15	1.03	1.14	1.13	1.03
ND ($\rho = 0.0$)	SS	1.98	2.55	1.01	2.12	2.49	1.02
	SS+SR/1	1.98	2.55	1.01	2.12	2.49	1.02
	SS+SR/2	1.98	2.55	1.01	2.12	2.49	1.02
Standard ($\rho = 0.0$)		1.98	2.54	1.00	2.12	2.49	1.02
Mixed ($\rho = \hat{\rho}$)		1.99	2.55	1.02	2.09	2.48	1.00

Table 8: $n = 100$: ERGM(3) network with fixed number of friends & $X \sim N(0, 25)$: Ratio of MSE(EBLUP) to MSE(BLUP)

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP - actual MSE		18,165	22,508	22,440	17,919	20,438	20,424
$\mathbf{Z}_U$ known ($\rho = 0.0$)		1.00	1.03	1.03	1.05	1.15	1.04
CN ($\rho = 0.0$)	SS	17.5	27.8	1.02	17.8	30.7	1.05
	SS+SR/1	2.55	5.61	1.05	2.67	5.91	1.05
	SS+SR/2	2.55	5.61	1.05	2.67	5.91	1.05
AR ($\rho = 0.0$)	SS	18.2	28.5	1.02	18.4	31.3	1.05
	SS+SR/1	2.77	5.37	1.05	2.89	5.59	1.05
	SS+SR/2	2.77	5.37	1.05	2.89	5.59	1.05
ND ($\rho = 0.0$)	SS	19.0	29.9	1.01	19.2	32.6	1.05
	SS+SR/1	18.7	28.4	1.02	19.0	31.1	1.03
	SS+SR/2	18.7	28.4	1.02	19.0	31.1	1.03
Standard ($\rho = 0.0$)		18.8	29.8	1.02	19.0	32.8	1.05
Mixed ($\rho = \hat{\rho}$)		19.0	30.3	1.03	19.3	33.4	1.04

Table 9: $n = 100$: ERGM(15) network & $X \sim N(0, 25)$: Ratio of average lengths of nominal 95% Gaussian CIs (EBLUP/BLUP), with % actual coverage in subscript

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP av(length)		$531_{95.9}$	$597_{95.6}$	$598_{95.3}$	$522_{95.2}$	$587_{94.8}$	$587_{95.0}$
$\mathbf{Z}_U$ known ($\rho = 0.0$)		$0.98_{94.5}$	$1.00_{94.0}$	$1.29_{95.0}$	$1.00_{94.9}$	$1.00_{93.3}$	$1.07_{94.0}$
CN ($\rho = 0.0$)	SS	$4.47_{94.0}$	$7.83_{93.5}$	$1.01_{95.3}$	$4.55_{93.4}$	$7.97_{93.9}$	$1.03_{94.1}$
	SS+SR/1	$0.98_{71.1}$	$5.35_{90.1}$	$1.01_{95.2}$	$1.00_{72.8}$	$5.45_{90.0}$	$1.02_{94.4}$
	SS+SR/2	$0.98_{71.9}$	$5.35_{90.2}$	$1.01_{95.2}$	$1.00_{73.2}$	$5.45_{90.1}$	$1.02_{94.3}$
AR ($\rho = 0.0$)	SS	$4.39_{92.8}$	$7.61_{91.7}$	$1.01_{94.9}$	$4.47_{91.9}$	$7.75_{92.8}$	$1.02_{94.0}$
	SS+SR/1	$2.32_{78.9}$	$1.68_{00.2}$	$1.01_{95.0}$	$2.36_{78.8}$	$1.73_{00.5}$	$1.03_{94.0}$
	SS+SR/2	$2.32_{79.0}$	$1.70_{00.4}$	$1.01_{95.0}$	$2.36_{79.0}$	$1.74_{00.6}$	$1.03_{94.0}$
ND ($\rho = 0.0$)	SS	$4.59_{93.6}$	$7.83_{92.9}$	$0.98_{94.5}$	$4.67_{93.0}$	$7.98_{93.8}$	$1.00_{93.5}$
	SS+SR/1	$4.30_{91.3}$	$7.91_{93.3}$	$0.98_{94.7}$	$4.39_{91.6}$	$8.05_{94.7}$	$1.00_{93.8}$
	SS+SR/2	$4.32_{91.3}$	$7.91_{93.4}$	$0.98_{94.6}$	$4.39_{91.8}$	$8.05_{94.7}$	$1.00_{93.8}$
Standard ($\rho = 0.0$)		$4.62_{94.1}$	$7.94_{93.7}$	$1.01_{95.3}$	$4.71_{93.0}$	$8.09_{94.4}$	$1.03_{94.4}$
Mixed ($\rho = \hat{\rho}$)		$4.65_{94.1}$	$7.97_{93.8}$	$1.02_{95.2}$	$4.73_{93.1}$	$8.12_{94.7}$	$1.02_{93.9}$

Table 10: $n = 100$: ERGM(15) network & $X \sim U\{1, .., 9\}$: Ratio of average lengths of nominal 95% Gaussian CIs (EBLUP/BLUP), with % actual coverage in subscript

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP av(length)		$531_{93.9}$	$540_{95.2}$	$540_{95.2}$	$522_{95.6}$	$531_{95.5}$	$531_{95.6}$
$\mathbf{Z}_U$ known ($\rho = 0.0$)		$0.98_{93.5}$	$0.98_{95.0}$	$0.98_{94.8}$	$0.997_{95.3}$	$0.998_{94.2}$	$0.997_{94.6}$
CN ($\rho = 0.0$)	SS	$1.27_{94.1}$	$1.45_{95.0}$	$0.99_{94.9}$	$1.30_{94.0}$	$1.47_{95.1}$	$1.01_{94.5}$
	SS+SR/1	$0.98_{92.4}$	$1.00_{93.4}$	$0.99_{94.9}$	$1.00_{93.9}$	$1.01_{92.8}$	$1.00_{94.3}$
	SS+SR/2	$0.98_{93.1}$	$1.00_{93.7}$	$0.99_{94.7}$	$1.00_{94.9}$	$1.01_{94.3}$	$1.01_{94.3}$
AR ($\rho = 0.0$)	SS	$1.27_{93.8}$	$1.44_{94.6}$	$0.99_{94.7}$	$1.29_{94.5}$	$1.46_{94.5}$	$1.00_{94.6}$
	SS+SR/1	$0.98_{92.5}$	$1.00_{93.1}$	$0.99_{94.6}$	$1.00_{93.6}$	$1.02_{93.4}$	$1.00_{94.2}$
	SS+SR/2	$0.97_{93.5}$	$0.99_{93.4}$	$0.99_{94.6}$	$0.99_{94.4}$	$1.01_{93.8}$	$1.00_{94.3}$
ND ($\rho = 0.0$)	SS	$1.29_{93.5}$	$1.48_{94.7}$	$0.98_{95.2}$	$1.32_{94.9}$	$1.50_{94.5}$	$1.00_{94.5}$
	SS+SR/1	$1.26_{92.4}$	$1.45_{94.2}$	$0.98_{94.6}$	$1.29_{93.9}$	$1.47_{93.6}$	$0.99_{95.0}$
	SS+SR/2	$1.27_{92.6}$	$1.46_{94.3}$	$0.98_{94.6}$	$1.30_{94.4}$	$1.48_{93.7}$	$1.00_{94.8}$
Standard ($\rho = 0.0$)		$1.30_{93.8}$	$1.49_{95.1}$	$0.99_{95.1}$	$1.33_{94.5}$	$1.51_{94.8}$	$1.01_{94.8}$
Mixed ($\rho = \hat{\rho}$)		$1.31_{93.7}$	$1.50_{95.1}$	$1.00_{95.1}$	$1.32_{93.9}$	$1.51_{94.8}$	$0.99_{94.4}$

Table 11: $n = 100$: ERGM(50) network & $X \sim N(0, 25)$: Ratio of average lengths of nominal 95% Gaussian CIs (EBLUP/BLUP), with % actual coverage in subscript

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP av(length)		$531_{95.8}$	$531_{95.1}$	$531_{95.2}$	$522_{94.9}$	$522_{93.9}$	$522_{93.8}$
$\mathbf{Z}_U$ known ($\rho = 0.0$)		$0.98_{94.6}$	$0.99_{94.3}$	$0.99_{93.9}$	$1.00_{94.3}$	$0.997_{93.0}$	$1.00_{92.7}$
CN ($\rho = 0.0$)	SS	$1.34_{95.0}$	$1.53_{95.1}$	$0.99_{94.1}$	$1.37_{94.1}$	$1.55_{93.5}$	$1.00_{93.2}$
	SS+SR/1	$0.98_{87.0}$	$0.99_{86.5}$	$0.99_{94.3}$	$1.00_{85.9}$	$1.01_{83.7}$	$1.00_{92.7}$
	SS+SR/2	$0.98_{94.1}$	$1.00_{94.1}$	$0.99_{94.4}$	$1.00_{93.6}$	$1.01_{92.0}$	$1.00_{92.9}$
AR ($\rho = 0.0$)	SS	$1.34_{94.9}$	$1.51_{94.8}$	$0.985_{94.1}$	$1.36_{94.0}$	$1.54_{93.3}$	$1.00_{92.9}$
	SS+SR/1	$1.00_{88.2}$	$1.02_{87.7}$	$0.99_{94.2}$	$1.02_{87.0}$	$1.03_{85.2}$	$1.00_{92.7}$
	SS+SR/2	$0.98_{94.0}$	$0.99_{93.9}$	$0.99_{94.0}$	$0.99_{93.5}$	$1.00_{91.3}$	$1.00_{92.6}$
ND ($\rho = 0.0$)	SS	$1.37_{94.9}$	$1.56_{94.5}$	$0.99_{94.1}$	$1.39_{93.6}$	$1.59_{93.8}$	$1.00_{92.7}$
	SS+SR/1	$1.35_{94.3}$	$1.55_{94.2}$	$0.98_{93.8}$	$1.37_{92.6}$	$1.58_{93.4}$	$1.00_{92.2}$
	SS+SR/2	$1.37_{94.7}$	$1.55_{94.5}$	$0.99_{93.9}$	$1.39_{93.3}$	$1.58_{93.5}$	$1.00_{92.5}$
Standard ($\rho = 0.0$)		$1.38_{95.0}$	$1.57_{94.4}$	$0.99_{94.5}$	$1.40_{93.5}$	$1.60_{93.7}$	$1.00_{93.1}$
Mixed ($\rho = \hat{\rho}$)		$1.38_{95.0}$	$1.57_{94.6}$	$1.00_{93.8}$	$1.40_{93.5}$	$1.60_{94.0}$	$0.99_{92.9}$

Table 12: $n = 100$: ERGM(50) & $X \sim U\{1, .., 9\}$: Ratio of average lengths of nominal 95% Gaussian CIs (EBLUP/BLUP), with % actual coverage in subscript

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP av(length)		$531_{94.0}$	$532_{95.1}$	$532_{95.0}$	$522_{95.2}$	$522_{94.9}$	$522_{94.9}$
$\mathbf{Z}_U$ known ($\rho = 0.0$)		$0.98_{93.3}$	$0.98_{94.7}$	$0.98_{94.5}$	$1.00_{95.6}$	$1.00_{94.5}$	$1.00_{94.9}$
CN ($\rho = 0.0$)	SS	$1.07_{92.9}$	$1.13_{95.6}$	$0.99_{94.7}$	$1.09_{94.6}$	$1.15_{94.9}$	$1.00_{94.3}$
	SS+SR/1	$0.98_{91.5}$	$0.99_{92.9}$	$0.98_{94.8}$	$1.00_{93.3}$	$1.00_{92.7}$	$1.00_{94.4}$
	SS+SR/2	$0.98_{93.2}$	$0.99_{93.5}$	$0.99_{94.5}$	$1.00_{95.6}$	$1.00_{94.1}$	$1.00_{94.2}$
AR ($\rho = 0.0$)	SS	$1.07_{92.9}$	$1.12_{95.7}$	$0.98_{94.5}$	$1.09_{94.7}$	$1.14_{95.1}$	$1.00_{94.8}$
	SS+SR/1	$1.00_{92.3}$	$1.01_{93.0}$	$0.98_{94.6}$	$1.01_{94.3}$	$1.03_{93.3}$	$1.00_{94.4}$
	SS+SR/2	$0.97_{93.2}$	$0.98_{94.1}$	$0.98_{94.4}$	$0.99_{95.3}$	$1.00_{94.5}$	$1.00_{94.4}$
ND ($\rho = 0.0$)	SS	$1.07_{93.2}$	$1.14_{95.5}$	$0.98_{94.6}$	$1.10_{95.0}$	$1.16_{94.6}$	$1.00_{94.8}$
	SS+SR/1	$1.06_{92.7}$	$1.13_{95.6}$	$0.98_{94.4}$	$1.08_{94.5}$	$1.15_{94.5}$	$0.99_{94.6}$
	SS+SR/2	$1.07_{93.1}$	$1.13_{95.5}$	$0.98_{94.5}$	$1.09_{94.7}$	$1.15_{94.7}$	$1.00_{94.9}$
Standard ($\rho = 0.0$)		$1.08_{93.4}$	$1.14_{95.4}$	$0.99_{94.6}$	$1.10_{94.9}$	$1.16_{95.0}$	$1.00_{94.5}$
Mixed ($\rho = \hat{\rho}$)		$1.08_{93.9}$	$1.15_{95.6}$	$0.99_{94.7}$	$1.09_{95.1}$	$1.15_{94.4}$	$0.99_{94.1}$

Table 13: $n = 100$: ERGM(3) network & $X \sim N(0, 25)$: Ratio of average lengths of nominal 95% Gaussian CIs (EBLUP/BLUP), with % actual coverage in subscript

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP av(length)		$531_{95.9}$	$600_{95.4}$	$600_{95.2}$	$521_{95.2}$	$589_{93.6}$	$589_{93.2}$
$\mathbf{Z}_U$ known ($\rho = 0.0$)		$0.98_{94.6}$	$1.00_{94.2}$	$1.26_{94.4}$	$1.00_{94.9}$	$1.01_{92.6}$	$1.24_{93.1}$
CN ($\rho = 0.0$)	SS	$4.59_{94.3}$	$6.05_{93.4}$	$1.01_{94.6}$	$4.67_{93.2}$	$6.17_{93.6}$	$1.03_{93.7}$
	SS+SR/1	$0.98_{72.1}$	$1.88_{82.4}$	$1.01_{94.9}$	$1.00_{71.9}$	$1.93_{84.2}$	$1.03_{93.1}$
	SS+SR/2	$0.98_{72.2}$	$1.88_{82.5}$	$1.01_{94.9}$	$1.00_{72.3}$	$1.93_{84.2}$	$1.03_{93.1}$
AR ($\rho = 0.0$)	SS	$4.51_{92.8}$	$5.88_{91.4}$	$1.00_{94.6}$	$4.59_{91.7}$	$5.99_{91.6}$	$1.02_{93.3}$
	SS+SR/1	$1.05_{74.6}$	$1.71_{80.5}$	$1.00_{94.9}$	$1.07_{73.7}$	$1.74_{79.5}$	$1.02_{93.0}$
	SS+SR/2	$1.06_{74.6}$	$1.71_{80.8}$	$1.00_{94.9}$	$1.07_{73.3}$	$1.75_{79.6}$	$1.02_{93.0}$
ND ($\rho = 0.0$)	SS	$4.71_{93.6}$	$6.13_{92.8}$	$0.99_{94.0}$	$4.80_{93.5}$	$6.25_{92.9}$	$1.00_{92.9}$
	SS+SR/1	$4.34_{92.3}$	$5.61_{92.3}$	$0.98_{94.6}$	$4.41_{90.8}$	$5.72_{94.1}$	$0.99_{92.7}$
	SS+SR/2	$4.35_{92.0}$	$5.62_{92.5}$	$0.98_{94.7}$	$4.42_{90.9}$	$5.73_{94.0}$	$1.00_{92.8}$
Standard ($\rho = 0.0$)		$4.74_{94.0}$	$6.24_{93.7}$	$1.01_{95.1}$	$4.83_{93.4}$	$6.36_{94.1}$	$1.03_{93.4}$
Mixed ($\rho = \hat{\rho}$)		$4.77_{94.0}$	$6.27_{93.6}$	$1.02_{94.8}$	$4.85_{93.5}$	$6.39_{93.9}$	$1.02_{93.0}$

Table 14: $n = 100$: ERGM(3) network & $X \sim U\{1, .., 9\}$: Ratio of average lengths of nominal 95% Gaussian CIs (EBLUP/BLUP), with % actual coverage in subscript

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP av(length)		$532_{94.3}$	$600_{95.7}$	$600_{95.8}$	$522_{95.1}$	$590_{95.2}$	$590_{95.3}$
$\mathbf{Z}_U$ known ($\rho = 0.0$)		$0.98_{93.3}$	$0.99_{95.3}$	$1.20_{95.3}$	$1.00_{95.2}$	$1.00_{93.4}$	$1.23_{93.7}$
CN ($\rho = 0.0$)	SS	$2.33_{92.5}$	$2.98_{94.6}$	$1.01_{94.8}$	$2.37_{93.3}$	$3.04_{93.9}$	$1.02_{94.1}$
	SS+SR/1	$0.98_{88.0}$	$1.26_{90.4}$	$1.01_{94.5}$	$0.99_{88.5}$	$1.28_{89.7}$	$1.02_{93.8}$
	SS+SR/2	$0.98_{88.5}$	$1.26_{90.3}$	$1.01_{94.5}$	$0.99_{88.5}$	$1.28_{89.7}$	$1.02_{93.8}$
AR ($\rho = 0.0$)	SS	$2.29_{91.4}$	$2.91_{93.2}$	$1.00_{94.5}$	$2.34_{92.1}$	$2.96_{92.4}$	$1.02_{94.0}$
	SS+SR/1	$0.99_{87.8}$	$1.19_{89.5}$	$1.00_{94.2}$	$1.01_{89.1}$	$1.21_{88.5}$	$1.02_{93.6}$
	SS+SR/2	$0.99_{87.9}$	$1.19_{89.8}$	$1.00_{94.2}$	$1.01_{88.8}$	$1.21_{88.4}$	$1.02_{93.6}$
ND ($\rho = 0.0$)	SS	$2.38_{92.2}$	$3.01_{93.9}$	$0.98_{94.6}$	$2.42_{93.4}$	$3.07_{92.9}$	$1.00_{93.5}$
	SS+SR/1	$2.20_{91.4}$	$2.78_{95.0}$	$0.98_{94.7}$	$2.24_{90.8}$	$2.83_{93.2}$	$0.99_{93.4}$
	SS+SR/2	$2.21_{91.6}$	$2.78_{95.2}$	$0.98_{94.8}$	$2.25_{91.2}$	$2.83_{93.2}$	$0.99_{93.5}$
Standard ($\rho = 0.0$)		$2.39_{93.0}$	$3.07_{95.2}$	$1.01_{94.5}$	$2.44_{93.8}$	$3.12_{93.2}$	$1.02_{94.0}$
Mixed ($\rho = \hat{\rho}$)		$2.40_{93.4}$	$3.08_{94.7}$	$1.02_{94.5}$	$2.44_{93.7}$	$3.14_{93.4}$	$1.01_{93.7}$

Table 15: $n = 100$: ERGM(3) network network with fixed number of friends & $X \sim N(0, 25)$: Ratio of average lengths of nominal 95% Gaussian CIs (EBLUP/BLUP), with % actual coverage in subscript

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP av(length)		$531_{95.7}$	$575_{95.3}$	$575_{95.2}$	$522_{94.8}$	$565_{95.2}$	$565_{95.4}$
$\mathbf{Z}_U$ known ($\rho = 0.0$)		$0.98_{94.9}$	$0.98_{93.5}$	$0.98_{93.1}$	$1.00_{94.6}$	$1.01_{94.8}$	$1.00_{94.6}$
CN ($\rho = 0.0$)	SS	$3.94_{93.4}$	$5.18_{94.0}$	$1.00_{94.9}$	$4.00_{93.0}$	$5.28_{94.2}$	$1.02_{94.5}$
	SS+SR/1	$0.98_{78.1}$	$1.82_{87.4}$	$1.00_{94.0}$	$1.00_{76.6}$	$1.85_{86.0}$	$1.02_{94.5}$
	SS+SR/2	$0.98_{78.1}$	$1.82_{87.4}$	$1.00_{94.0}$	$1.00_{76.6}$	$1.85_{86.0}$	$1.02_{94.5}$
AR ($\rho = 0.0$)	SS	$3.91_{92.7}$	$5.08_{93.0}$	$0.99_{94.8}$	$3.97_{92.4}$	$5.18_{92.5}$	$1.01_{94.4}$
	SS+SR/1	$1.06_{78.7}$	$1.71_{86.1}$	$0.99_{94.0}$	$1.08_{78.1}$	$1.73_{84.8}$	$1.01_{94.4}$
	SS+SR/2	$1.06_{78.7}$	$1.71_{86.1}$	$0.99_{94.0}$	$1.08_{78.1}$	$1.73_{84.8}$	$1.01_{94.4}$
ND ($\rho = 0.0$)	SS	$4.07_{93.4}$	$5.30_{93.0}$	$0.98_{94.3}$	$4.14_{93.4}$	$5.40_{93.7}$	$1.00_{94.1}$
	SS+SR/1	$3.92_{92.2}$	$5.08_{92.8}$	$0.97_{93.2}$	$3.99_{92.1}$	$5.18_{93.6}$	$0.99_{94.2}$
	SS+SR/2	$3.92_{92.2}$	$5.08_{92.8}$	$0.97_{93.2}$	$3.99_{92.1}$	$5.18_{93.6}$	$0.99_{94.2}$
Standard ($\rho = 0.0$)		$4.12_{93.3}$	$5.40_{94.3}$	$1.00_{94.9}$	$4.19_{93.6}$	$5.50_{94.3}$	$1.02_{94.7}$
Mixed ($\rho = \hat{\rho}$)		$4.14_{93.4}$	$5.43_{94.3}$	$1.00_{94.7}$	$4.21_{93.4}$	$5.53_{94.2}$	$1.01_{94.7}$

Table 16: $n = 100$: ERGM(3) network network with fixed number of friends & $X \sim U\{1, .., 9\}$: Ratio of average lengths of nominal 95% Gaussian CIs (EBLUP/BLUP), with % actual coverage in subscript

		Population Data Generated Under Model					
		$\rho = 0.0$			$\rho = 0.1$		
EBLUP Based On		CN	AR	ND	CN	AR	ND
BLUP av(length)		$532_{94.5}$	$575_{94.5}$	$576_{94.5}$	$522_{95.2}$	$565_{94.7}$	$565_{94.5}$
$\mathbf{Z}_U$ known ($\rho = 0.0$)		$0.98_{93.3}$	$0.99_{93.7}$	$0.98_{93.6}$	$1.00_{95.3}$	$1.01_{92.6}$	$1.00_{92.7}$
CN ($\rho = 0.0$)	SS	$2.04_{93.5}$	$2.59_{94.7}$	$1.0_{94.4}$	$2.08_{94.0}$	$2.64_{94.6}$	$1.02_{93.3}$
	SS+SR/1	$0.98_{88.5}$	$1.23_{90.0}$	$1.00_{94.8}$	$1.00_{90.6}$	$1.25_{90.3}$	$1.02_{92.7}$
	SS+SR/2	$0.98_{88.5}$	$1.23_{90.0}$	$1.00_{94.8}$	$1.00_{90.6}$	$1.25_{90.3}$	$1.02_{92.7}$
AR ($\rho = 0.0$)	SS	$2.03_{93.2}$	$2.55_{93.8}$	$0.99_{94.3}$	$2.07_{92.9}$	$2.59_{94.0}$	$1.01_{93.0}$
	SS+SR/1	$0.99_{88.5}$	$1.18_{88.6}$	$1.00_{94.5}$	$1.01_{89.9}$	$1.20_{89.3}$	$1.01_{92.7}$
	SS+SR/2	$0.99_{88.5}$	$1.18_{88.6}$	$1.00_{94.5}$	$1.01_{89.9}$	$1.20_{89.3}$	$1.01_{92.7}$
ND ($\rho = 0.0$)	SS	$2.10_{93.1}$	$2.64_{94.3}$	$0.98_{93.6}$	$2.14_{93.8}$	$2.69_{94.0}$	$1.00_{93.1}$
	SS+SR/1	$2.03_{91.0}$	$2.55_{93.8}$	$0.98_{93.7}$	$2.07_{92.2}$	$2.59_{93.1}$	$0.99_{92.8}$
	SS+SR/2	$2.03_{91.0}$	$2.55_{93.8}$	$0.98_{93.7}$	$2.07_{92.2}$	$2.59_{93.1}$	$0.99_{92.8}$
Standard ($\rho = 0.0$)		$2.12_{94.2}$	$2.69_{94.8}$	$1.00_{94.7}$	$2.16_{94.6}$	$2.74_{94.7}$	$1.02_{93.2}$
Mixed ($\rho = \hat{\rho}$)		$2.13_{94.4}$	$2.71_{94.8}$	$1.01_{94.8}$	$2.17_{94.7}$	$2.75_{94.4}$	$1.01_{93.4}$