Informative prior on structural equation modelling with non-homogenous error structure [version 2; peer review: 2 approved]

Oladapo A. Olalude, Bernard O. Muse, Oluwayemisi O. Alaba

1Department of Statistics, University of Ibadan, Ibadan, Oyo State, +234, Nigeria
2Department of Mathematics and Statistics, Rufus Giwa Polytechnic, Owo, Ondo State, Nigeria

Abstract

Introduction: This study investigates the impact of informative prior on Bayesian structural equation model (BSEM) with heteroscedastic error structure. A major drawback of homogeneous error structure is that, in most studies the underlying assumption of equal variance across observation is often unrealistic, hence the need to consider the non-homogenous error structure.

Methods: Updating appropriate informative prior, four different forms of heteroscedastic error structures were considered at sample sizes 50, 100, 200 and 500.

Results: The results show that both posterior predictive probability (PPP) and log likelihood are influenced by the sample size and the prior information, hence the model with the linear form of error structure is the best.

Conclusions: The study has been able to address sufficiently the problem of heteroscedasticity of known form using four different heteroscedastic conditions, the linear form outperformed other forms of heteroscedastic error structure thus can accommodate any form of data that violates the homogenous variance assumption by updating appropriate informative prior. Thus, this approach provides an alternative approach to the existing classical method which depends solely on the sample information.

Keywords
Bayesian SEM, Latent Variable, Observed Variable, Heteroscedastic error structure, Predictive Performance

Open Peer Review

Approval Status ✓ ✓

version 2 (revision)
20 Sep 2022

version 1
04 May 2022

1. Adenike Oluwafunmilola Olubiyi, Ekiti State University, Ado Ekiti, Nigeria
2. Mohamed R. Abonazel, Cairo University, Giza, Egypt

Any reports and responses or comments on the article can be found at the end of the article.
Introduction

Bayesian structural equation modeling (BSEM) analyzes the relationship between the observed, unobserved, and latent variables within the Bayesian context. In spite the rising number of statistical research ideas that have been created and verified using structural equation modeling (SEM). Despite its propensity to skew statistical estimates and inference and unlike the classical regression, we suggest the use of diagnostic tests for the presence of multicollinearity, heteroscedasticity, and nonnormality. Bayesian structural equation model investigations rarely mention the use of statistical approaches for measurement and structural model assessment, non-normality, multicollinearity, heteroscedasticity, and combinations thereof.

In Bayesian inference, \( \theta \) is random, which depicts the level of uncertainty about the true value of \( \theta \) because both the observed data \( y \) and the parameters \( \theta \) are assumed random. The joint probability of the parameters and the data as functions of the conditional distribution of the data given the parameters, and the prior distribution of the parameters can be modelled. More formally,

\[
p(\theta|Y) \propto p(\theta)p(Y|\theta)
\]

where

- \( p(\theta|Y) \) is the posterior distribution
- \( p(\theta) \) is the prior distribution
- \( p(Y|\theta) \) is the likelihood function

The un-normalized posterior distribution when expressed in terms of the unknown parameters \( \theta \) for fixed values of \( y \), this term is the likelihood \( L(\theta|y) \). Thus, can be rewritten as:

\[
p(\theta|Y) \propto p(\theta)L(\theta|y)
\]

Studies abound on classical methods and Bayesian methods with a focus on homogeneous variance. This study explores the BSEM using different forms of heteroscedastic error structure.

Methods

Bayesian estimation of structural equation models (SEM)

This section develops a Gibbs sampler to estimate SEM with reflective measurement indicators. The Bayesian estimation is illustrated by considering a SEM that is equivalent to the mostly used model. A SEM is composed of a measurement equation (3) and a structural equation (4):

\[
y_i = \Lambda \omega_i + \epsilon_i
\]

\[
\eta_i = \Pi \eta_i + \Gamma \xi_i + \delta_i
\]

where \( i \in \{1, \ldots, n\} \).
It is assumed that measurement errors are uncorrelated with \( \omega \) and \( \delta \), residuals are uncorrelated with \( \omega \) and the variables are distributed as follows:

\[
\epsilon_i \sim N(0, \Psi_\epsilon)
\]

(5)

\[
\delta_i \sim N(0, \Psi_\delta)
\]

(6)

\[
\omega_i \sim N(0, \Sigma_\omega)
\]

(7)

\( \forall i \in \{1, \ldots, n\} \), where \( \Psi_\epsilon \) and \( \Psi_\delta \) are diagonal matrices. The covariance matrix of \( \omega \) is derived based on the SEM:

\[
\Sigma_\omega = \begin{bmatrix}
E(\eta^T) & E(\xi^T) \\
E(\eta^T) & E(\xi^T)
\end{bmatrix}
\]

(8)

\[
\Sigma_\omega = \begin{bmatrix}
\Pi_0^{-1}(\Gamma \Phi \Gamma^T + \Psi_\delta) \Pi_0^{-T} & \Pi_0^{-1} \Gamma \\
\Phi \Gamma^T \Pi_0^{-T} & \Phi
\end{bmatrix}
\]

(9)

\[
\eta^T = (\Pi_0^{-1} \Gamma \xi + \Pi_0^{-1} \delta) \Pi_0^{-T} \\
= \Pi_0^{-1} (\Gamma \xi^T \Gamma^T + \delta \xi^T) \Pi_0^{-T} + \Pi_0^{-1} (\Gamma \xi^T \delta^T + \delta \xi^T \Gamma^T) \Pi_0^{-T}
\]

(10)

\[
\eta^T = E(\eta^T) = \Pi_0^{-1} (\Gamma \Phi \Gamma^T + \Psi_\delta) \Pi_0^{-T} \\
\xi^T = (\Pi_0^{-1} \Gamma \xi + \Pi_0^{-1} \delta) \xi^T \\
E(\xi^T) = \Pi_0^{-1} \Gamma
\]

Prior distributions

In order to enable Gibbs sampling from full conditional posterior distributions, natural conjugate prior distributions for the unknown parameters are considered. 26 Let \( \psi_{\kappa} \) be the \( k \)th diagonal element of \( \Psi_\epsilon \), \( \psi_i \) be the \( l \)th diagonal element of \( \Psi_\delta \), \( \Lambda_{\kappa} \) be the \( k \)th row of \( \Lambda \) and \( M_{\kappa} \) be the \( l \)th row of \( M \),

\[
\psi_{\kappa}^{-1} \sim \text{Gamma}(a_{0\kappa}, b_{0\kappa})
\]

(11)

\[
[\Lambda_{\kappa} | \psi_{\kappa}^{-1}] \sim N(\Lambda_{0\kappa}, \psi_{\kappa}^{-1}, H_{0\kappa})
\]

(12)

\[
\psi_i^{-1} \sim \text{Gamma}(a_{0i}, b_{0i})
\]

(13)

\[
[M_i | \psi_i^{-1}] \sim N(M_{0i}, \psi_i^{-1}, H_{0i})
\]

(14)

\[
\Phi \sim \text{IW}(V_0, \nu_0)
\]

(15)

with \( \kappa \in \{1, \ldots, p\} \) and \( i \in \{1, \ldots, q\} \)

Derivations of conditional distributions

The joint posterior of all unknown parameters is proportional to the likelihood times the prior,

\[
p(\Lambda, \Psi_\epsilon, \Omega, M, \Psi_\delta, \Phi | Y) \propto p(Y | \Lambda, \Psi_\epsilon, \Omega, M, \Psi_\delta, \Phi) \ast p(\Lambda, \Omega, M, \Psi_\delta, \Phi)
\]

(16)
Given \(Y\) and \(\Omega\), \(A\) and \(\Psi\), are independent from \(\Sigma\). Draws of \(\Omega\), can cause estimation of \(A\) and \(\Psi\) as a simple regression model. Thus, sampling from the posterior distribution of \(A\) and \(\Psi\), sampling is employed. This involves the use of marginal posterior distribution. Since the full posterior distribution is intractable; a Markov chain Monte Carlo (MCMC) simulation method of Gibbs with regard to \(M, \Phi, \Psi\) and \(\Sigma\), which are independent from \(Y\) given \(\Omega\).

**Heteroscedastic error structures**

The heteroscedastic error structure with different functional form of error variance under consideration are double logarithmic form, linear form, linear-inverse form and linear-absolute form as expressed inequation 17, 18, 19 and 20, respectively.

\[
\sigma^2 = \ln \sigma^2 = \lambda_0^* + \lambda_1^* \ln \gamma^*_l + vi
\]

(17)

\[
\sigma^2 = |\epsilon^*_l e^*_l| = \left(\lambda_0^* + \lambda_2^* \gamma^*_l + vi\right)^2
\]

(18)

\[
\sigma^2 = |\epsilon^*_l e^*_l| = \left(\lambda_0^* + \lambda_2^* \sqrt{\gamma^*_l} + vi\right)^2
\]

(19)

\[
\sigma^2 = |\epsilon^*_l e^*_l| = \left(\lambda_0^* + \lambda_2^* |\gamma^*_l| + vi\right)^2
\]

(20)

Each of the functional forms of heteroscedastic error structure will be incorporated into the modified model. The variance matrix for disturbance vector is given as

\[
\sum = \left(\begin{array}{cc}
\sigma^2 \lambda_1^* & 0 \\
0 & \sigma^2 \lambda_2^*
\end{array}\right)
\]

(21)

\[
\Omega = \left[\begin{array}{ccc}
\sigma^2 \lambda_1^* & 0 & \cdots & 0 \\
0 & \sigma^2 \lambda_2^* & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma^2 \lambda_2^*
\end{array}\right]
\]

(22)

**The posterior distribution**

The posterior density is the product of the likelihood and the prior distribution chosen.

\[
(P(\lambda^*, h, \Omega) | y^*) \exp (y^* | \lambda^*, h, \Omega)p(\lambda^*)p(h)p(\Omega)
\]

(23)

\[
p(P(\lambda^*, h), \Omega|y^*) = h^N 2^{N/2} |\Omega|^{-1/2} (2\pi)^{-N/2} \exp \left[-\frac{1}{2} \left(y^* - \lambda_0^* \right)^T \Omega^{-1} \left(y^* - \lambda_0^* \right) \right] \times \exp \left[-\frac{1}{2} \left(\lambda^* - \lambda_0^* \right)^T \Omega^{-1} \left(\lambda^* - \lambda_0^* \right) \right] \times \exp \left[-\frac{1}{2} \left(\lambda^* - \lambda_0^* \right)^T \Omega^{-1} \left(\lambda^* - \lambda_0^* \right) \right] 
\]

\[
\times \exp \left[-\frac{1}{2} \left(\lambda^* - \lambda_0^* \right)^T \Omega^{-1} \left(\lambda^* - \lambda_0^* \right) \right] 
\]

(24)

\[
P(\lambda^*, h, \Omega|y^*) = h^N 2^{N/2} |\Omega|^{-1/2} (2\pi)^{-N/2} \exp \left[-\frac{1}{2} \left(y^* - \lambda_0^* \right)^T \Omega^{-1} \left(y^* - \lambda_0^* \right) \right] \times \exp \left[-\frac{1}{2} \left(\lambda^* - \lambda_0^* \right)^T \Omega^{-1} \left(\lambda^* - \lambda_0^* \right) \right] 
\]

\[
\times \exp \left[-\frac{1}{2} \left(\lambda^* - \lambda_0^* \right)^T \Omega^{-1} \left(\lambda^* - \lambda_0^* \right) \right] 
\]

(25)

Since the full posterior distribution is intractable; a Markov chain Monte Carlo (MCMC) simulation method of Gibbs sampling is employed. This involves the use of marginal posterior distribution.

\[
\lambda = \lambda_0^* = (y^* \Omega^{-1} \gamma)^{-1} y^* \Omega^{-1} \gamma = (y^* \gamma)^{-1} y^* \gamma
\]

(26)

\[
S^2 = \left(\frac{y^* - \gamma \lambda_0^*}{\sum} \right) \gamma^* \gamma^* \left(\lambda^* - \lambda_0^* \right) = (y^* - \gamma \lambda^*)' (y^* - \gamma \lambda^*)
\]

Also

\[
\mathbf{V}_S^2 + \left(\lambda^* - \lambda_0^* \right)' \gamma^* \gamma^* \left(\lambda^* - \lambda_0^* \right) = (y^* - \gamma \lambda^*)' (y^* - \gamma \lambda^*)
\]
\[
p(\gamma^* | \eta, \sigma^2) = \frac{h^{\gamma^*/2}}{(2\pi)^{k/2}} \exp \left( -\frac{h}{2} \left( \frac{1}{2}S^2 + (\lambda^* - \lambda_0)^\prime \gamma^\ast \gamma^\ast (\lambda^* - \lambda_0) \right) \right)
\]

where \(\nu = N - K\) and \(N = \nu + K\).

Consider an informative prior created by set.

\[
\nu^{-1} j = \left( \frac{1}{c^j} \right) \gamma^\ast_j
\]

And letting \(c \to 0\) for \(j = 1, 2\).

The posterior distribution of \(\lambda^*\) conditional on \(\gamma^*, h, \Omega\) is given by:

\[
p(\lambda^* | \gamma^*, h, \Omega) \propto h^k \exp \left( -\frac{h}{2} (\gamma^* - \lambda^*)' \Omega^{-1} (\gamma^* - \lambda^*) + (\lambda^* - \lambda_0)' \nu^{-1} (\gamma^* - \lambda_0) \right)
\]

\[
\times \exp \left[ -\frac{h}{2(\Omega)} (\gamma^* - \lambda^*)' (\gamma^* - \lambda^*) + \frac{(\lambda^* - \lambda_0)' (\lambda^* - \lambda_0)}{\nu} \right]
\]

Solving the exponential part of the above equation, we will have:

\[
(\gamma^* - \lambda^*)' (\gamma^* - \lambda^*) = (\gamma^*)^2 + (\lambda^*)^2 - 2\gamma^* \lambda^* + (\gamma^* - \lambda_0)' (\gamma^* - \lambda_0) = (\lambda^*)^2 + \left( \lambda_0' - 2 \lambda^* \right)^2
\]

Therefore,

\[
= \exp \left[ -\frac{h}{2(\Omega)} \sum_{i=1}^{N} \gamma_{i}^2 + (\lambda^*)^2 - 2\gamma^* \lambda^* + \frac{(\lambda^* - \lambda_0)^2}{\nu} \right]
\]

The additional term not involving \(\lambda^*\) is factored out to give:

\[
= \exp \left[ -\frac{\lambda^2}{2\nu} + \frac{\lambda^2 \lambda_0^2}{2\nu} - \frac{n\lambda^2}{2\nu \Omega^2} \right]
\]

Factorization in terms of \(\lambda^*\), the term in the exponential becomes:

\[
= -\frac{\lambda^2}{\sigma^2} + \frac{2\lambda^2 \lambda_0^2}{2\sigma^2}
\]

\[
\sigma^2 = \left( \frac{1}{\nu} + \frac{n}{\Omega \nu \sigma^2} \right)^{-1} \text{ and } \lambda^* = \sigma^2 \left( \lambda_0^2 \nu + \frac{n\lambda^2}{\Omega \nu \sigma^2} \right)
\]

So, the posterior density of \(\lambda^*\) conditioned on other parameter \(h, \Omega, \gamma^*\) is a multivariate normal with mean \(\lambda^*\) and variance \(\sigma^2\).

That is,

\[
p(\lambda^* | h, \Omega, \gamma^*) \sim N(\lambda^*, \sigma^2)
\]

The posterior distribution of \(h\) conditional on \(\lambda^*, \Omega, \gamma^*\) is given by:

\[
P(h | \lambda^*, \Omega, \gamma^*) \propto h^{N-2k} \exp \left[ -\frac{h}{2} (\gamma^* - \lambda^*)' \Omega^{-1} (\gamma^* - \lambda^*) \times h^k \exp \left( -\frac{hv}{2s^2} \right) \right]
\]

\[
= h^{N-2k} \exp \left[ -\frac{h}{2k} \sum_{i=1}^{N} \left( \gamma_{i}^2 + n(\lambda^*)^2 - 2\gamma^* \lambda^* \right) - \frac{hv}{2s^2} \right]
\]
The posterior distribution of $\Omega^*$, conditional on $y^*, \lambda^*, h$, is given by:

$$
P(\Omega^*|y^*, \lambda^*, h) \propto P(\Omega) \times h^N \exp \left[ -\frac{h}{2} (y^* - \lambda^* y^\top)^\top \Omega^{-1} (y^* - \lambda^* y^\top) \right]
$$

(31)

$$
P(\Omega^*|y^*, \lambda^*, h) \propto h^N \exp \left[ -\frac{h}{2} (y^* - \lambda^* y^\top)^\top \Omega^{-1} (y^* - \lambda^* y^\top) \right] \times |\Omega^*|^{-(\rho_0 + k + 1)/2} \exp \text{tr}(R_0^{-1}\Sigma_0^{-1})/2
$$

(32)

**The Gibbs sampler**

The Gibbs sampling procedure used in this study involves generation of sequence of draws from the conditional posterior distribution of each parameter.2,22,26

**Gibbs sampling procedure**

(i) Choose a starting or initial value, $\phi^{(0)}$ for $s = 1, 2, \ldots, S$

(ii) Take a random draw, $\phi_1^{(s)}$ from the full conditional, $p(\phi_1|y, \phi^{(s-1)}_1)$

(iii) Take a random draw, $\phi_2^{(s)}$ from the full conditional, $p(\phi_2|y, \phi^{(s)}_1)$ using the updated values of $\phi_1^{(s)}$

(iv) Repeat until $M$ draws are obtained, each being a vector of $\phi^{(s)}$

(v) Perform the Burn-in by dropping the first $S_{0}$ of these draws to eliminate the effect of $\phi_0$, the remaining $S_{1}$ draws are then averaged to obtain the estimate of the posterior $E[g(\phi)|y]$.

The right-hand side of (15) is proportional to the density function of an inverse Wishart distribution. Then,

$$
P(\Phi|Y, \Omega) \sim \text{IW}_q \left[ (\Omega\Omega^\top R_0^{-1}), n + \rho_0 \right]
$$

(33)

**Design of simulation**

- At different functional forms of heteroscedastic error structure with changes in sample size of 50, 100, 200 and 500. Hyper-parameter will be arbitrarily chosen for the simulation using Gibbs sampler an MCMC method.6,22

- The R code can be accessed via the Extended data.27

- Factor loading and error precision followed multivariate normal and inverse gamma distributions respectively to assess the prior sensitivity.21

- The criteria that will be used to assess the performance of the posterior simulation technique are the posterior estimates.

In order to evaluate the Bayesian model fit, we used the posterior predictive probability (PPP) procedure.4,5,7,25

$$
\text{PPP} = P \left( y, \lambda^* | y, \lambda^* \right) < f \left( y^{\text{rep}}, \lambda^{\text{rep}} \right) \equiv \frac{1}{m} \sum_{i=1}^{m} \delta_i
$$

(34)

After achieving convergence (after $j$ iterations), $\left( \lambda^{(j+1)}, \Omega^{(j+1)} \right)$ can be regarded as observation from $p(\lambda^*, \Omega|y)$ collect $\left( \lambda^{(t)}, \Omega^{(t)} \right)$ $t = j + 1, \ldots, + T$ for statistical inference.

$$
\lambda = T^{-1} \sum_{t=1}^{T} \lambda^{(t)}, \Omega = T^{-1} \sum_{t=1}^{T} \Omega^{(t)}
$$

(35)

gives Bayesian estimates of parameter and the latent variables.10,17,23
Results and discussion
The section presents the discussion of analysis of results; performances of the estimators across the parameters for the different forms of heteroscedasticity, performances of Bayesian posterior simulation and analytical methods in the presence of heteroscedasticity via consideration of four (4) different forms of heteroscedastic error structures over four sample sizes of 50, 100, 200 and 500.

Performance of the estimators at heteroscedasticity condition
This gives the results for the latent and observed variables at various sample sizes for the four heteroscedastic error conditions considered.

Comparison of latent variable estimates at different sample sizes under the heteroscedasticity condition
Using the assumed values for the estimates which are $\lambda_1 = 2.0$, $\lambda_2 = 3.0$ and precision = 15.0.

The covariance matrix of $\omega$ was derived to be $E \cdot \eta \xi^T = \prod_0^1 \Gamma$ with $M$ at fixed values (0 or 1). The Bayesian estimates of SEM using the independent normal-gamma priors were derived for the two classes of SEM. Hyper-parameter was arbitrarily chosen for the simulation using Gibbs sampler a Markov chain Monte Carlo (MCMC) method since the joint posterior density does not have a tractable form. For the double logarithmic form, at 95% credible interval, when $n=50$, Posterior Mean, PM, and Precision, PR (2.011, 2.435, and 13.202), Posterior Standard Deviation PSD (0.035, 0.033, and 0.223) and when $n=100$, PM, and PR (2.022, 2.528, and 13.70), PSD (0.023, 0.025, and 0.251), when $n=200$, PM, and PR (2.052, 2.611, and 14.4), PSD (0.017, 0.018, and 0.255), when $n=500$, PM, and PR (2.010, 2.801, and 14.7), PSD (0.031, 0.021, and 0.258).

For the linear form, when $n=50$, PM, and PR (1.845, 2.779, and 13.95), PSD (0.240, 0.242, and 0.235). When $n=100$, PM, and PR (1.861, 2.811, and 14.22), PSD (0.328, 0.226, and 0.325), when $n=200$, PM, and PR (1.956, 2.921, and 14.72), PSD (0.219, 0.217, and 0.212), and when $n=500$, PM, and PR (2.109, 2.801, and 14.95), PSD (0.031, 0.021, and 0.258).

For the linear-inverse form when $n=50$, PM, and PR (1.882, 2.742, and 14.95), PSD (0.040, 0.028, and 0.291). When $n=100$, PM, and PR (1.972, 2.835, and 14.65), PSD (0.024, 0.023, and 0.229). When $n=200$, PM, and PR (1.988, 2.901, and 14.45), PSD (0.017, 0.016, and 0.109), and when $n=500$, PM, and PR (2.021, 3.003, and 14.21), PSD (0.011, 0.015, and 0.105).

For the linear-absolute form, when $n=50$, PM, and PR (2.036, 2.824, and 14.50), PSD (0.032, 0.034, and 0.122). When $n=100$, PM, and PR (1.908, 2.903, and 13.92), PSD (0.022, 0.026, and 0.234). When $n=200$, PM, and PR (1.893, 2.809, and 13.85), PSD (0.017, 0.023, and 0.311), and when $n=500$, PM, and PR (1.806, 2.788, and 13.55), PSD (0.031, 0.035, and 0.433).

Table 1. Double logarithmic form on latent variable and observed variable estimates.

| Sample sizes | Latent variables | Posterior Mean (PM) | Posterior Standard Deviation (PSD) | Credible Interval (CI) | Measured variables | Estimate | Standard Deviation |
|--------------|------------------|---------------------|-----------------------------------|------------------------|-------------------|---------|-------------------|
| n=50         | $\lambda_1$     | 2.011               | 0.035                             | 1.959                  | $x_1$             | 0.045   | 0.023             |
|              | $\lambda_2$     | 2.435               | 0.033                             | 2.384                  | $x_2$             | 0.038   | 0.023             |
|              | Precision (PR)  | 13.202              | 0.223                             | 13.071                 |                  |         |                   |
| n=100        | $\lambda_1$     | 2.022               | 0.023                             | 1.979                  | $x_1$             | 0.053   | 0.008             |
|              | $\lambda_2$     | 2.528               | 0.025                             | 2.484                  | $x_2$             | 0.037   | 0.024             |
|              | Precision (PR)  | 13.700              | 0.251                             | 13.561                 |                  |         |                   |
| n=200        | $\lambda_1$     | 2.052               | 0.017                             | 2.015                  | $x_1$             | 0.006   | 0.045             |
|              | $\lambda_2$     | 2.611               | 0.018                             | 2.573                  | $x_2$             | 0.048   | 0.020             |
|              | Precision (PR)  | 14.4                | 0.255                             | 14.260                 |                  |         |                   |
| n=500        | $\lambda_1$     | 2.010               | 0.031                             | 1.961                  | $x_1$             | 0.040   | 0.028             |
|              | $\lambda_2$     | 2.801               | 0.021                             | 2.760                  | $x_2$             | 0.018   | 0.004             |
|              | Precision (PR)  | 14.7                | 0.258                             | 14.559                 |                  |         |                   |
Examining different forms of heteroscedastic error structures in Bayesian structural equation modeling using informative priors, rather than assuming homogenous variance which is often a statistical fallacy in many studies. We compare the models’ posterior means and standard deviations in Tables 1, 2, 3 and 4. The differences are unlikely to impact substantive conclusions, but two of them are noteworthy.

First, the posterior means of the loadings ( $\lambda_1$ and $\lambda_2$ ) are somewhat smaller under different heteroscedastic condition with the informative priors as observed in Tables 6 and 7. Second, the factor variance $\gamma^*$ is larger under our model with informative priors, likely because the informative prior placed more density on larger values of the posterior standard deviation. An evaluation of the model fit was based on the values of PPP as shown in Table 5 and it was observed that the linear form is the best with minimum PPP value as sample size increases. It was also revealed by the downward slope of the model as the sample size increases from 50 to 500 shown in Figure 1b when compared with Figure 1a, 2a and 2b.

### Table 2. Linear form on latent variable and observed variable estimates.

| Sample sizes | Latent variables | Posterior Mean (PM) | Posterior Standard Deviation (PSD) | Credible Interval (CI) | Measured variables | Estimate | Standard Deviation |
|--------------|------------------|---------------------|------------------------------------|------------------------|-------------------|----------|--------------------|
| n=50         | $\lambda_1$      | 1.845               | 0.240                              | 1.709                   | $x_1$             | 0.078    | 0.017              |
|              | $\lambda_2$      | 2.779               | 0.242                              | 2.643                   | $x_2$             | 0.055    | 0.036              |
|              | Precision         | 13.950              | 0.235                              | 13.816                  | 14.844            |          |                    |
| n=100        | $\lambda_1$      | 1.861               | 0.328                              | 1.702                   | $x_1$             | 0.079    | 0.012              |
|              | $\lambda_2$      | 2.811               | 0.226                              | 2.679                   | $x_2$             | 0.036    | 0.028              |
|              | Precision         | 14.220              | 0.325                              | 14.062                  | 14.378            |          |                    |
| N=200        | $\lambda_1$      | 1.956               | 0.219                              | 1.826                   | $x_1$             | 0.071    | 0.008              |
|              | $\lambda_2$      | 2.921               | 0.217                              | 2.792                   | $x_2$             | 0.047    | 0.016              |
|              | Precision         | 14.72               | 0.212                              | 14.542                  | 14.898            |          |                    |
| N=500        | $\lambda_1$      | 2.120               | 0.211                              | 1.993                   | $x_1$             | 0.052    | 0.022              |
|              | $\lambda_2$      | 3.122               | 0.311                              | 2.967                   | $x_2$             | 0.059    | 0.010              |
|              | Precision         | 14.95               | 0.114                              | 14.857                  | 15.044            |          |                    |

### Table 3. Linear inverse form on latent variable and observed variable estimates.

| Sample sizes | Latent variables | Posterior Mean (PM) | Posterior Standard Deviation (PSD) | Credible Interval (CI) | Measured variables | Estimate | Standard Deviation |
|--------------|------------------|---------------------|------------------------------------|------------------------|-------------------|----------|--------------------|
| n=50         | $\lambda_1$      | 1.882               | 0.043                              | 1.827                   | $x_1$             | 0.075    | 0.020              |
|              | $\lambda_2$      | 2.742               | 0.028                              | 2.696                   | $x_2$             | 0.023    | 0.017              |
|              | Precision         | 14.95               | 0.291                              | 14.801                  | 15.099            |          |                    |
| n=100        | $\lambda_1$      | 1.972               | 0.024                              | 1.929                   | $x_1$             | 0.055    | 0.010              |
|              | $\lambda_2$      | 2.835               | 0.023                              | 2.793                   | $x_2$             | 0.031    | 0.021              |
|              | Precision         | 14.65               | 0.229                              | 14.317                  | 14.583            |          |                    |
| N=200        | $\lambda_1$      | 1.988               | 0.017                              | 1.826                   | $x_1$             | 0.054    | 0.006              |
|              | $\lambda_2$      | 2.901               | 0.016                              | 2.790                   | $x_2$             | 0.032    | 0.024              |
|              | Precision         | 14.45               | 0.109                              | 14.358                  | 14.541            |          |                    |
| N=500        | $\lambda_1$      | 2.021               | 0.011                              | 1.992                   | $x_1$             | 0.052    | 0.015              |
|              | $\lambda_2$      | 3.003               | 0.015                              | 2.969                   | $x_2$             | 0.050    | 0.022              |
|              | Precision         | 14.210              | 0.105                              | 14.120                  | 14.300            |          |                    |
### Table 4. Linear absolute form on latent variable and observed variable estimates.

| Sample sizes | Latent variables | Posterior Mean (PM) | Posterior Standard Deviation (PSD) | Credible Interval (CI) | Measured variables | Estimate | Standard Deviation |
|--------------|------------------|---------------------|-----------------------------------|------------------------|--------------------|----------|-------------------|
| n=50         | λ₁               | 2.036               | 0.032                             | 1.986 - 2.086          | x₁                 | 0.043    | 0.018             |
|              | λ₂               | 2.824               | 0.034                             | 2.773 - 2.875          | x₂                 | 0.027    | 0.022             |
|              | Precision        | 14.500              | 0.122                             | 14.403 - 14.597        |                    |          |                   |
| n=100        | λ₁               | 1.908               | 0.022                             | 1.867 - 1.949          | x₁                 | 0.047    | 0.017             |
|              | λ₂               | 2.903               | 0.026                             | 2.858 - 2.948          | x₂                 | 0.043    | 0.025             |
|              | Precision        | 13.92               | 0.234                             | 13.786 - 14.054        |                    |          |                   |
| N=200        | λ₁               | 1.893               | 0.017                             | 1.857 - 1.929          | x₁                 | 0.054    | 0.017             |
|              | λ₂               | 2.809               | 0.023                             | 2.767 - 2.851          | x₂                 | 0.041    | 0.024             |
|              | Precision        | 13.85               | 0.311                             | 13.696 - 14.005        |                    |          |                   |
| N=500        | λ₁               | 1.806               | 0.031                             | 1.757 - 1.855          | x₁                 | 0.048    | 0.019             |
|              | λ₂               | 2.788               | 0.035                             | 2.736 - 2.840          | x₂                 | 0.044    | 0.022             |
|              | Precision        | 13.55               | 0.433                             | 13.367 - 13.732        |                    |          |                   |

Note: Posterior mean (PM), posterior standard deviation (PSD), credible interval (CI).

### Table 5. Comparison at varying sample sizes of different heteroscedastic form.

| Sample size | Double logarithmic | Linear | Linear inverse | Linear absolute |
|-------------|--------------------|--------|----------------|-----------------|
|             | LogLik PPP         | LogLik PPP | LogLik PPP | LogLik PPP | LogLik PPP | LogLik PPP | LogLik PPP |
| N=50        | -17.577 0.538      | -17.309 0.501 | -19.701 0.567 | -20.065 0.560 |
| N=100       | -24.324 0.543      | -43.058 0.523 | -16.214 0.544 | -19.777 0.544 |
| N=200       | -29.427 0.541      | -44.935 0.545 | -15.305 0.540 | -19.547 0.532 |
| N=500       | -35.510 0.482      | -60.920 0.570 | -14.494 0.531 | -18.171 0.506 |

### Table 6. Latent variable estimates at different sample sizes under the double-logarithmic and linear forms.

| Sample size | Latent variables | Double logarithmic | Linear |
|-------------|------------------|--------------------|--------|
|             |                  | Posterior Mean (PM) | Posterior Standard Deviation (PSD) | Credible Interval (CI) | Posterior Mean (PM) | Posterior Standard Deviation (PSD) | Credible Interval (CI) |
| N=50        | λ₁               | 2.001              | 0.231  | 1.868 - 2.134 | 2.110 | 0.230  | 1.977 - 2.243 |
|             | λ₂               | 2.283              | 0.538  | 2.080 - 2.486 | 2.554 | 0.201  | 2.430 - 2.678 |
| N=100       | λ₁               | 2.021              | 0.312  | 1.866 - 2.176 | 2.020 | 0.123  | 1.923 - 2.117 |
|             | λ₂               | 2.478              | 0.562  | 2.270 - 2.686 | 2.601 | 0.356  | 2.436 - 2.766 |
| N=200       | λ₁               | 2.032              | 0.432  | 1.850 - 2.214 | 2.011 | 0.174  | 1.895 - 2.127 |
|             | λ₂               | 2.770              | 0.832  | 2.517 - 3.023 | 2.705 | 0.456  | 2.518 - 2.892 |
| N=500       | λ₁               | 2.100              | 0.445  | 1.915 - 2.285 | 2.005 | 0.253  | 1.866 - 2.144 |
|             | λ₂               | 2.888              | 1.564  | 2.541 - 3.234 | 3.102 | 0.575  | 2.892 - 3.312 |

Note: Posterior mean (PM), posterior standard deviation (PSD), credible interval (CI).
Table 7. Latent variable estimates at different sample sizes under the linear-inverse and linear absolute forms.

| Sample size | Latent variables | Linear-inverse | Linear-inverse | Linear-inverse | Linear-inverse | Linear-inverse | Linear-inverse | Linear-inverse | Linear-inverse |
|-------------|------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|             | Posterior Mean (PM) | Posterior Standard Deviation (PSD) | Credible Interval (CI) | Posterior Mean (PM) | Posterior Standard Deviation (PSD) | Credible Interval (CI) | Posterior Mean (PM) | Posterior Standard Deviation (PSD) | Credible Interval (CI) | Posterior Mean (PM) | Posterior Standard Deviation (PSD) | Credible Interval (CI) |
| N=50        | λ₁              | 2.101          | 0.352          | 1.937          | 2.265          | 1.732          | 0.311          | 1.577          | 1.887          | 1.732          | 0.311          | 1.577          | 1.887          |
|             | λ₂              | 2.637          | 0.528          | 2.436          | 2.838          | 2.582          | 0.583          | 2.370          | 2.794          | 2.582          | 0.583          | 2.370          | 2.794          |
| N=100       | λ₁              | 1.982          | 0.421          | 1.802          | 2.162          | 1.810          | 0.252          | 1.671          | 1.949          | 1.810          | 0.252          | 1.671          | 1.949          |
|             | λ₂              | 2.754          | 0.192          | 2.633          | 2.875          | 2.634          | 0.375          | 2.464          | 2.804          | 2.634          | 0.375          | 2.464          | 2.804          |
| N=200       | λ₁              | 1.975          | 0.476          | 1.784          | 2.166          | 1.820          | 0.211          | 1.696          | 1.947          | 1.820          | 0.211          | 1.696          | 1.947          |
|             | λ₂              | 2.814          | 0.901          | 2.551          | 3.077          | 2.723          | 0.766          | 2.480          | 2.966          | 2.723          | 0.766          | 2.480          | 2.966          |
| N=500       | λ₁              | 2.111          | 0.488          | 1.917          | 2.305          | 1.920          | 0.145          | 1.815          | 2.026          | 1.920          | 0.145          | 1.815          | 2.026          |
|             | λ₂              | 3.073          | 1.102          | 2.782          | 3.364          | 2.902          | 0.331          | 2.743          | 3.062          | 2.902          | 0.331          | 2.743          | 3.062          |

Figure 1. Plot of log likelihood and posterior predictive probability (PPP) at various sample sizes under (a) the double logarithmic form and (b) the linear form.

Figure 2. Plot of log likelihood and posterior predictive distribution (PPP) at various sample sizes under (a) the linear-inverse form (b) the linear-absolute form.
Considering an improvement to maximum likelihood method, in Bayesian estimations, parameters are considered as random with informative prior distribution also known as the conjugate family of the posterior, once the data is simulated/collected, it is combined with prior distribution using Bayes theorem, next posterior distribution is calculated reflecting the prior knowledge and simulated data.14,15,21 Joint posterior distribution is summarized using MCMC simulation techniques in terms of lower dimensional summary statistics as posterior mean and posterior standard deviations.2,26 We observe that the structural and measurement equation obtained from this study are adequate and in general we could accept the proposed model.

Conclusion
In this research, the derived Bayesian estimators of a structural equation model in the presence of different forms of heteroscedastic error structures validated accurate statistical inference. The study has also been able to address sufficiently the problem of heteroscedasticity of known form using four different heteroscedastic conditions for both linear and quadratic forms, and it has also successfully modified the homogenous error structure to heteroscedastic error structure in Bayesian structural equation model.20 The linear form outperformed other forms of heteroscedastic error structure thus can accommodate any form of data that violates the homogenous variance assumption by updating appropriate informative prior. However, these heteroscedastic error structure models can also be tested as an area of further research by updating appropriate noninformative prior.16,18 Thus, this approach provides an alternative approach to the existing classical method which depends solely on the sample information.

Data availability

Underlying data
All data underlying the results are available as part of the article and no additional source data are required.

Extended data

Figshare: RCODE BSEM.docx. https://doi.org/10.6084/m9.figshare.19299851.26

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

References

1. Anderson JC, Gerbing DW: Some methods for respecifying measurement models to obtain unidimensional construct measurement. Journal of Marketing Research. 1982; 19(4): 453–460. Publisher Full Text
2. Ansari A, Jedidi K: Bayesian Factor Analysis for Multilevel Binary Observations. Psychometrika. 2000; 65(4): 475–496. Publisher Full Text
3. Ansari A, Jedidi K, Dubé L: Heterogeneous factor analysis models: A Bayesian approach. Psychometrika. 2002; 67(1): 49–77. Publisher Full Text
4. Asparouhov T, Muthén BO: Bayesian analysis of latent variable models using Mplus. 2010. Reference Source
5. Bansal S: A new Gibbs sampling-based Bayesian model updating approach using modal data from multiple setups. International Journal for Uncertainty Quantification. 2015; 5(4): 361–374. Publisher Full Text
6. Bellman R: Dynamic programming and Lagrange multipliers. Proceedings of the National Academy of Sciences of the United States of America. 1956; 42(10): 767–769. Publisher Full Text | Published Abstract | Free Full Text
7. Bentler PM: Comparative Fit Indexes in Structural Models. Psychological Bulletin. 1990; 107(2): 238–246. PubMed Abstract | Publisher Full Text
8. Das S, Chen M-H, Kim S, et al.: A Bayesian Structural Equations Model for Multilevel Data with Missing Responses and Missing Covariates. 2008; vol 3, Number 1, pp. 197–224
9. Depaoli S: Measurement and structural model class separation in mixture CFA: ML/EM versus MCMC. Structural Equation Modeling. 2012; 19: 178–203. Publisher Full Text
10. Dunson DB: Bayesian Latent Variable Models for Clustered Mixed Outcomes. Journal of the Royal Statistical Society, Series B, 2002; 62: 355–366. Publisher Full Text
11. Hancock GR, Mueller RO: Structural equation modeling: A second course. Information Age Publishing, Inc.; 2006.
12. Kaplan D, Depaoli S: Bayesian statistical methods. In T. D. Little (Ed.), Oxford handbook of quantitative methods. Oxford: Oxford University Press; 2013 (pp. 407–437).
13. Kass RE, Raftery AE: Bayes Factors. Journal of the American Statistical Association. 1995; 90: 773–795. Publisher Full Text
14. Lee S, Song X: Maximum Likelihood Analysis of a General Latent Variable Model with Hierarchically Mixed Data. Biometrics. 2003; 60: 624–636.
15. Lee S, Song X: Maximum Likelihood Analysis of a General Latent Variable Model with Hierarchically Mixed Data. Biometrics. 2003; 60: 624–636.
16. Lee SY, Shi JQ: Bayesian analysis of structural equation model with fixed covariates.. Structural Equation Modeling. 2000; 7: 411–430. Publisher Full Text
17. Mauricio G, Jorgensen TD: Adapting Fit Indices for Bayesian Structural Equation Modeling. Comparison to Maximum Likelihood Journal of Psychological Methods. 2019; 25(1): 46–70. Publisher Full Text
18. Meghan K, Cai1 and Zhang Z: Fit for a Bayesian: An Evaluation of PPP and DIC for Structural Equation Modeling 2018; 26: 39–50. Publisher Full Text
19. Oberski DL, Satorra A: Measurement error models with uncertainty about the error variance. Structural Equation Modeling. 2013; 20(3): 409–428. Publisher Full Text
20. Olsson UH, Foss T, Troye SV, et al.: The performance of ML, GLS, and WLS estimation in SEM under conditions of misspecification and non-normality. Structural Equation Modeling. 2000; 2000; 7: 557–595.

21. Palomo J, Dunson DB, Bollen K: Bayesian structural equation modeling. Handbook of Computing and Statistics with Application. 2007; 1: 163–188. Publisher Full Text

22. Scheines R, Hoijtink H, Boomsma A: Bayesian estimation and testing of structural equation models. Psychometrika. 1999; 64: 37–52. Publisher Full Text

23. Spiegelhalter DJ, et al.: Bayesian measures of model complexity and fit. Journal of Royal Statistical Society B. 2002; 64(4): 583–639.

24. Sanghee L, Nigel PM: Robustness of structural equation modeling to distributional misspecification: empirical evidence & research guidelines.

25. Van Erp S, Mulder J, Oberski DL: Prior sensitivity analysis in default Bayesian structural equation modeling. Psychological Methods. 2016; 21(2): 363–388. PubMed Abstract | Publisher Full Text

26. Yanuar F, Ibrahim K, Abdul AJ: Bayesian structural equation modeling for the health index. Journal of Applied Statistics. 2013; 40(6): 1254–1269. Publisher Full Text

27. Olalude O, Alaba OO, Muse O, et al.: RCODE BSEM.docx, Tables and the figures. figshare. Dataset. 2022. Publisher Full Text
Open Peer Review

Current Peer Review Status:

✓ ✓

Version 2

Reviewer Report 27 September 2022

https://doi.org/10.5256/f1000research.137309.r150953

© 2022 Abonazel M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mohamed R. Abonazel

Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt

I am happy with the corrections in the revised paper. It was improved. So, the current version of this manuscript is suitable for indexing. Good luck.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Econometrics, R Programming, Panel Data, Time Series, Computational Statistics, Data Analysis, R Statistical Packages, Statistical Modeling, Nonparametric Models, Robust Regression.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 26 July 2022

https://doi.org/10.5256/f1000research.120326.r136840

© 2022 Abonazel M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mohamed R. Abonazel

Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
This paper investigated the impact of informative prior on Bayesian structural equation model with heteroscedastic error structure. Four different forms of heteroscedastic error structures were considered. The suggested Bayesian approach provides an alternative approach to the existing classical method which depends solely on the sample information. The results indicate that the suggested Bayesian estimation method is more efficient than the existing classical method.

In my opinion, the paper offers a good contribution. So, I recommend accepting this paper, but after making the following modifications to improve the manuscript:

1. I think the title of the paper needs improvement. I suggest the following title: “Bayesian estimation of structural equation modelling with non-homogenous error structure”.

2. In the “abstract” section, the findings or research results should be introduced briefly in the abstract.

3. In the “introduction” section, the introduction did not contain enough background information. Also discuss the similar work that has been done in this area to give a detailed view of this work. The authors should add more papers related to the Bayesian estimation of structural equation modelling.

4. In the “methods” section, the authors should define each symbol given in each equation.

5. In the “conclusion” section, the limitation and future research directions should be mentioned.

Is the work clearly and accurately presented and does it cite the current literature?  
Partly

Is the study design appropriate and is the work technically sound?  
Yes

Are sufficient details of methods and analysis provided to allow replication by others?  
Yes

If applicable, is the statistical analysis and its interpretation appropriate?  
Yes

Are all the source data underlying the results available to ensure full reproducibility?  
Yes

Are the conclusions drawn adequately supported by the results?  
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Applied statistics, Econometric models.
I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Adenike Oluwafunmilola Olubiyi
Department of Statistics, Ekiti State University, Ado Ekiti, Nigeria

ABSTRACT
1. In this paper, the research team investigates the impact of informative prior on Bayesian Structural equation model (BSEM) with heteroscedastic error structure.

2. The drawback of homogeneous error structure was addressed by considering the non-homogenous error structure.

General Comment under this section: the researcher can add their findings

INTRODUCTION
1. Bayesian structural equation modelling (BSEM) analyses the relationship between the observed, unobserved and latent variables within the Bayesian context

2. The likelihood is the un-normalized posterior distribution when expressed in terms of the unknown parameters $\theta$ for fixed values of $y$.

3. The study explores the BSEM using different forms of heteroscedastic error structure.

4. Other studies abound on classical methods and Bayesian methods with focus on homogeneous variance.\cite{8,19,22,25}

General Comment under this section: The Introductory part was well presented and detailed with relevant citation, even though the researchers can still explore more.

Methods
1. Gibbs sampler was developed to estimate SEM with reflective measurement indictors.\cite{1,11,12}

2. The SEM equation used is composed of a measurement equation and a structural equation.

3. To enable Gibbs sampling from full conditioner posterior distributions, natural conjugate prior distributions for the unknown parameters were considered.

4. The heteroscedastic error structure with different functional form of error variance under
consideration are double logarithm form, linear form, linear-inverse form and linear-absolute form as expresses in equation 17,18,19 and 20.

5. Markov Chain Monte Carlo (MCMC) Simulation method of Gibbs Sampling was employed.

SIMULATION
1. At different functional forms of heteroscedastic error structure with changes in sample size of 50,100,200 and 500. Hyper-parameter was arbitrarily chosen for the simulation using Gibbs sampler an MCMC method.

2. To assess the prior sensitivity, factor loading and error precision followed multivariate normal and inverse gamma distributions respectively.

3. The posterior estimate is used to assess the performance of the posterior simulation technique.

4. In order to evaluate the Bayesian Model fit, the researcher used the posterior predictive probability (PPP) procedure.4,5,7,24

General Comment under this section: The methods were well presented and the simulation study well organized.

RESULTS
1. This gives the results for the latent and observed variables at various sample sizes for the four heteroscedastic error conditions considered using the assumed values for the estimates which are $\lambda_1 = 2.0, \lambda_2 = 3.0$ and precision 15.0.

2. The posterior means of loadings $\lambda_1$ and $\lambda_2$ are somewhat smaller under different heteroscedastic condition with the informative priors.

3. It was observed that the linear form is the best with minimum PPP value as sample size increases.

4. It was also revealed by the downward slope of the model as the sample size increases from 50 500.

5. It was observed that the structural and measurement equation obtained from this study are adequate and in general could be accepted for the proposed model.

General comment under this section: The obtained results in this section indicate the correct performances with increased sample sizes and the incorporation of informative priors.

CONCLUSION
1. The study has been able to address sufficiently the problem of heteroscedasticity of known form using four different heteroscedasticity of known form using four different heteroscedastic conditions for both linear and quadratic forms.

2. It has also successfully modified the homogenous error structure to heteroscedastic error
structure in Bayesian structural equation model.

3. Thus, the approach provides an alternative approach to the existing classical method which depends solely on sample information.

**General comment:** this section flows with the contents of the paper and is well presented. The manuscript is well written and followed the format of the journal and has substance; the manuscript can be approved.

**References**
1. Ansari A, Jedidi K: Bayesian factor analysis for multilevel binary observations. *Psychometrika*. 2000; **65** (4): 475-496 Publisher Full Text
2. Anderson J, Gerbing D: Some Methods for Respecifying Measurement Models to Obtain Unidimensional Construct Measurement. *Journal of Marketing Research*. 1982; **19** (4). Publisher Full Text
3. Bellman R: DYNAMIC PROGRAMMING AND LAGRANGE MULTIPLIERS. *Proc Natl Acad Sci U S A*. 1956; **42** (10): 767-9 PubMed Abstract | Publisher Full Text

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Yes

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Environmental Statistics and Econometric

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
The benefits of publishing with F1000Research:

• Your article is published within days, with no editorial bias
• You can publish traditional articles, null/negative results, case reports, data notes and more
• The peer review process is transparent and collaborative
• Your article is indexed in PubMed after passing peer review
• Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com