Study 1: Comparative Performance of Gaussian and Skew-Normal Copula VAR Models Under Marginal Misspecification

Table of contents

0. Summary

0.1 Computational Stability versus Statistical Inference

The Normal–Gaussian (NG) model is computationally stable across all simulation conditions based on $\hat{R}$ (max $\hat{R} \leq 1.01$ in all replications). The Skew–Gaussian (SG) model becomes harder to sample as marginal skewness increases: the proportion of replications with $\hat{R} > 1.01$ rises under extremeCHI (roughly one-quarter to one-third of runs, depending on T). Divergent transitions, when recorded, can be used to refine this classification.

0.2 Model Performance Across Skewness Conditions

Moderate skewness (moderateSN, $\alpha = \pm 4$): NG and SG exhibit comparable performance. For $T \geq 100$, bias is near zero and coverage is close to nominal on average for the core parameters. At $T = 50$, both models show mild finite sample attenuation in some Φ and ρ parameters (due to priors, estimation uncertainty or errors in variables effect?), but the magnitudes are small and do not affect coverage. Under this level of skewness, SG's additional marginal flexibility yields no inferential benefit.

Strong skewness (strongSN, $\alpha = \pm 9$): Differences between NG and SG emerge for the VAR dynamics. SG reduces relative bias in Φ modestly (on average ≈ 0.04, up to ≈ 0.12 in some cells) and converges faster to the true values as T increases. Both models exhibit downward bias in ρ. For the NG model, this arises from marginal misspecification (assuming normal when the truth is skew-normal). For the SG model, despite being correctly specified in terms of the distributional family, the shape parameter α is difficult to estimate at small T (see Section 7.2): when $\hat{\alpha}$ is biased toward zero, the fitted CDF still deviates from the truth, producing PIT distortion and attenuated ρ estimates. As T increases and α becomes better identified, this distortion diminishes.

Extreme misspecification (extremeCHI, standardized χ_1^2 innovations): Performance deteriorates markedly, with parameter-specific failure modes:

- VAR dynamics (Φ): SG shows smaller bias and conservative (over-nominal) coverage for $T \geq 100$, whereas NG is closer to nominal coverage on average but shows larger bias and occasional undercoverage in the worst cells.
- Copula correlation (ρ): both models exhibit severe attenuation; in a minority of replications (across directions) posterior means can cross zero, producing relative bias below -1.
- Intercepts (μ): SG shows systematic bias in μ (absolute bias up to ≈ 0.36 in some settings), while NG's μ remains close to 0. This appears to reflect an interaction between the centered skew-normal parameterization and the inability of the skew-normal family to approximate the χ_1^2 tail behavior.

0.3 Insights

PIT distortion induced by marginal CDF misspecification is the dominant failure mechanism. In both models, misspecified marginals yield non-uniform PITs, which attenuate the effective dependence seen by the Gaussian copula and can lead to sign reversals in a minority of replications under extreme mismatch. The NG model can partially accommodate skewness through small scale adjustments (typically $|\sigma - 1| \approx 0.02$–$0.03$), but this does not correct tail probabilities and therefore does not repair the PIT. Under extremeCHI, the SG model's intercept

estimates shift away from 0, suggesting that the intercept absorbs part of the marginal mis-fit under the centered skew-normal parameterization; this empirical trade-off coincides with relatively stable estimation of Φ but does not remedy attenuation of ρ.

1. Introduction

This simulation study compares the performance of two Bayesian Vector Autoregressive (VAR(1)) models: a standard Normal-Gaussian (NG) model and a Skew-Gaussian (SG) model. The study investigates how these models recover the true parameters when the data generating process (DGP) exhibits varying degrees of skewness in the innovations, coupled by a Gaussian copula.

1.1. Data Generating Process (DGP)

The DGP is a bivariate VAR(1) model:

$$Y_t = \mu + \Phi Y_{t-1} + \epsilon_t$$

Where Y_t is a 2x1 vector of observations, μ is the intercept vector (set to 0 in this simulation), Φ is the 2x2 matrix of autoregressive coefficients, and ϵ_t is the 2x1 vector of innovations (errors).

The innovations $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})^T$ are generated such that they are standardized.

> **i** Standardization of the innovations
>
> We require innovations $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})^\top$ to have $\mathbb{E}[\epsilon_{i,t}] = 0$ and $\mathrm{Var}(\epsilon_{i,t}) = 1$ for all conditions. This is enforced in the DGP, and the fitted models are parameterized to be consistent with this convention.
>
> **Skew-Normal marginals (DGP and SG model)** Let $\mathrm{SN}(\xi, \omega, \alpha)$ denote the skew-normal with location ξ, scale $\omega > 0$, shape α.
>
> $$\delta = \frac{\alpha}{\sqrt{1 + \alpha^2}}$$
>
> Moments:
>
> $$\mu_{\mathrm{SN}} = \xi + \omega\delta\sqrt{\frac{2}{\pi}}, \quad \sigma_{\mathrm{SN}}^2 = \omega^2 \left(1 - \frac{2\delta^2}{\pi}\right)$$
>
> To impose mean 0 and variance 1, choose

$$\omega = \left(1 - \frac{2\delta^2}{\pi}\right)^{-\frac{1}{2}}, \quad \xi = -\omega\delta\sqrt{\frac{2}{\pi}}$$

In the DGP (for moderateSN and strongSN), we set $\alpha \in \{\pm 4, \pm 9\}$, compute δ, then pick ω, ξ via the formulas above so that draws from $\mathrm{SN}(\xi, \omega, \alpha)$ are standardized.

In the SG model, we use centered parameterization that re-expresses the SN so that the innovation has zero theoretical mean by construction. In practice, the Stan parameters are (δ, ω) (or equivalent reparameterizations), and the implied ξ is that value which centers the innovation at 0 (as above). This is why the SG model cannot use μ to absorb a nonzero innovation mean; μ therefore represents the VAR intercept (process mean) rather than a marginal location parameter.

Chi-squared marginals (DGP and NG/SG models) For χ^2_ν: $\mathbb{E} = \nu$, $\mathrm{Var} = 2\nu$, skewness $\gamma_1 = \sqrt{8/\nu}$, excess kurtosis $\gamma_2 = 12/\nu$. In the DGP (for extremeCHI), we standardize after simulating: if $X \sim \chi^2_1$, set

$$Z = \frac{X - 1}{\sqrt{2}},$$

and (optionally) mirror as $-Z$ for left skew; in the copula DGP this is implemented via $Q_{-X}(u) = -Q_X(1 - u)$ so the copula parameter is preserved.

Thus Z has mean 0 and variance 1 but remains skewed ($\gamma_1 \approx 2.83$) and heavy-tailed ($\gamma_2 = 12$). In the NG model, innovations are Gaussian with scale σ. Because the data are standardized, the truth is $\sigma = 1$; deviations from 1 arise from misspecification.

In the SG model, innovations remain skew-normal with zero-mean by construction, using (δ, ω). Under a χ^2 DGP this implies misspecified marginals; the parameter α adjusts shape flexibly but cannot match χ^2_1 skewness.

ℹ SG Model Parameterization: Variance Not Fixed

The SG model estimates both scale (ω) and shape (δ) parameters freely, rather than constraining the innovation variance to equal 1. This apparent overparameterization is acceptable for several reasons:

1. **Bayesian identification:** In a Bayesian framework with proper priors on both ω and δ, the posterior remains well-defined even without fixing variance. The data inform both parameters jointly, and the priors regularize the solution.

2. **Practical flexibility:** Fixing $\omega = f(\delta)$ to enforce unit variance would reduce model flexibility and could cause numerical issues when $|\delta| \to 1$ (where $\omega \to \infty$). The free parameterization avoids these boundary complications.

3. **Prior-induced regularization:** The half-normal prior on ω (centered near 1) and the normal prior on δ (centered at 0) together induce a prior on innovation variance that concentrates around reasonable values without hard constraints.

4. **Robustness to misspecification:** When the true DGP is not skew-normal (e.g., extremeCHI), the model can adjust both shape and scale to best approximate the data, rather than being forced into a potentially poor fit by the unit-variance constraint.

The joint distribution of ϵ_t is modeled using a Gaussian Copula, parameterized by the correlation coefficient ρ. This allows the dependence structure to be modeled independently of the marginal distributions $f_i(\epsilon_{i,t})$. The joint density is given by Sklar's theorem:

$$f(\epsilon_{1,t}, \epsilon_{2,t}) = c(u_{1,t}, u_{2,t}; \rho) \cdot f_1(\epsilon_{1,t}) \cdot f_2(\epsilon_{2,t})$$

$$f(\epsilon_{1,t}, \epsilon_{2,t}) = c_{\text{Gauss}}(F_1(\epsilon_{1,t}), F_2(\epsilon_{2,t}); \rho) \cdot \frac{2}{\omega_1} \phi\left(\frac{\epsilon_{1,t} - \xi_1}{\omega_1}\right) \Phi\left(\alpha_1 \frac{\epsilon_{1,t} - \xi_1}{\omega_1}\right) \cdot \frac{2}{\omega_2} \phi\left(\frac{\epsilon_{2,t} - \xi_2}{\omega_2}\right) \Phi\left(\alpha_2 \frac{\epsilon_{2,t} - \xi_2}{\omega_2}\right)$$

The marginal distributions f_i are varied to introduce different levels and types of skewness.

> **!** Copula Sign Under Mirroring (Preserved via Q_{-X})
>
> Left-skew margins are produced by mirroring standardized innovations. In the DGP we implement mirroring via the quantile function of $-X$,
>
> $$Q_{-X}(u) = -Q_X(1 - u),$$
>
> which is monotone increasing in u. This preserves the Gaussian copula parameter ρ even in mixed-direction cells.
> **Implementation in this report.** Bias and coverage summaries for the copula parameter use the input ρ directly; no sign adjustment is applied. In the key simulation plots, the *input* $\rho \in \{0.30, 0.50\}$ is encoded by **line type** (rather than an additional facet dimension) to keep panels readable.
> **Nuance.** A naive mirroring $-Q_X(u)$ would flip the copula sign and would require an effective-ρ adjustment. We avoid that by using $1 - u$ in the quantile step.

> **i** Numerical Stability in Copula Evaluation
>
> The Gaussian copula density requires evaluating $\Phi^{-1}(u)$ where $u = F(\epsilon)$ is the probability integral transform. When u approaches 0 or 1, $\Phi^{-1}(u)$ diverges to $\pm\infty$, causing numerical overflow. To prevent this, we apply boundary clamping:
>
> $$u_{\text{clamped}} = \max(\varepsilon, \min(1 - \varepsilon, u)), \quad \varepsilon = 10^{-9}$$
>
> This clamping affects only the most extreme quantiles (beyond the 0.9999999th percentile)

and has negligible impact on inference. The choice of $\varepsilon = 10^{-9}$ balances numerical stability against loss of tail information—values smaller than 10^{-12} risk floating-point underflow, while values larger than 10^{-6} would noticeably compress the effective copula support.

1.2. Simulation Design

The study employs a full factorial design crossing five factors, resulting in 108 unique conditions, with 200 replications per condition.

Table 1: Summary of the Simulation Design Factors.

Factor	Levels
Time Series Length (T)	50, 100, 200
Copula Correlation (ρ)	0.30, 0.50
VAR Parameters (Φ)	**Set A (Symmetric)**: $\begin{pmatrix} 0.40 & 0.10 \\ 0.10 & 0.40 \end{pmatrix}$
	Set B (Asymmetric): $\begin{pmatrix} 0.55 & 0.10 \\ 0.10 & 0.25 \end{pmatrix}$
Skewness Level (Marginals)	**moderateSN**: Skew-Normal (SN), shape $\alpha = \pm 4$
	strongSN: SN, shape $\alpha = \pm 9$
	extremeCHI: Standardized Chi-squared (χ_1^2)
Skewness Direction	++ (both positive), -- (both negative), +- (mixed; -+ omitted)

> **i** Exclusion of -+ Direction
>
> The design includes directions ++, --, and +-, and omits -+ to reduce computational cost.
>
> - For VAR Set A ($\phi_{11} = \phi_{22}$ and $\phi_{12} = \phi_{21}$), relabeling $Y_1 \leftrightarrow Y_2$ leaves the DGP invariant. Under this symmetry, +- and -+ are exactly equivalent, and including both would be redundant.
>
> - For VAR Set B, the cross-effects remain symmetric ($\phi_{12} = \phi_{21}$), but the diagonal dynamics differ ($\phi_{11} \neq \phi_{22}$). Swapping Y_1 and Y_2 therefore also swaps ϕ_{11} and ϕ_{22}; -+ is not strictly identical to +- within the same VAR-Set-B condition.
>
> Accordingly, mixed-direction results for Set B should be interpreted qualitatively (e.g., patterns in PIT distortion and ρ attenuation) rather than as an exact surrogate for the

7

omitted `-+` case.

Recommendations. 1. If exact exchangeability under mixed skewness is required, restrict mixed-direction analyses to VAR Set A. 2. If VAR Set B is substantively important, include the `-+` direction (and, if needed, the swapped-diagonal analogue of Set B) as a sensitivity analysis.

1.3. True Parameter Values

Table 2: True Parameter Values Used in the Data Generating Process.

Parameter	True Value	Notes
μ_1, μ_2	0, 0	Innovations are mean-zero
ϕ_{11} (Set A / Set B)	0.40 / 0.55	Diagonal AR coefficients
$\phi_{12} = \phi_{21}$	0.10	Cross-effects (symmetric)
ϕ_{22} (Set A / Set B)	0.40 / 0.25	Diagonal AR coefficients
ρ	0.30 or 0.50	Copula correlation
σ_1, σ_2 (NG model)	1.0, 1.0	Innovations are unit-variance
α (moderateSN)	± 4 (direction-dependent)	$\delta \approx \pm 0.970$
α (strongSN)	± 9 (direction-dependent)	$\delta \approx \pm 0.994$
ω (moderateSN)	1.5795	Derived from $\alpha = \pm 4$
ω (strongSN)	1.6415	Derived from $\alpha = \pm 9$
α, ω (extremeCHI)	Not applicable (misspecified)	SG parameters have no true counterpart

> **i Interpreting SG Parameters Under extremeCHI**
>
> For the extremeCHI condition, the DGP uses standardized χ_1^2 marginals, not skew-normal. Therefore, the SG model parameters (α, ω) have no "true" values in the conventional sense—any estimate represents the model's best skew-normal approximation to a non-skew-normal distribution.
>
> In the analysis, we set these truth values to NA and exclude them from bias/coverage calculations. When interpreting SG model fits under extremeCHI, the estimated α and ω should be understood as:
>
> - α: The skewness direction and magnitude that best approximates χ_1^2 within the skew-normal family (expected to be large and positive for right-skewed conditions)
> - ω: The scale adjustment needed to match variance given the fitted shape
>
> These estimates reflect approximation quality rather than parameter recovery.

> **i** Bias Metric for Intercepts (μ)
>
> The "Relative Bias" plots display `mean_rel_bias`, defined as:
>
> $$\text{Relative Bias} = \frac{\hat{\theta} - \theta_{\text{true}}}{|\theta_{\text{true}}|}$$
>
> However, for the intercept parameters μ_1 and μ_2, the true value is zero, making relative bias undefined. In these cases, we report **absolute bias** instead:
>
> $$\text{Bias}_{\mu} = \hat{\mu} - 0 = \hat{\mu}$$
>
> This means that for μ panels in "Relative Bias" plots, the y-axis shows absolute deviation from zero (in original units), not a proportion. Cross-parameter comparisons should account for this difference in scale interpretation.
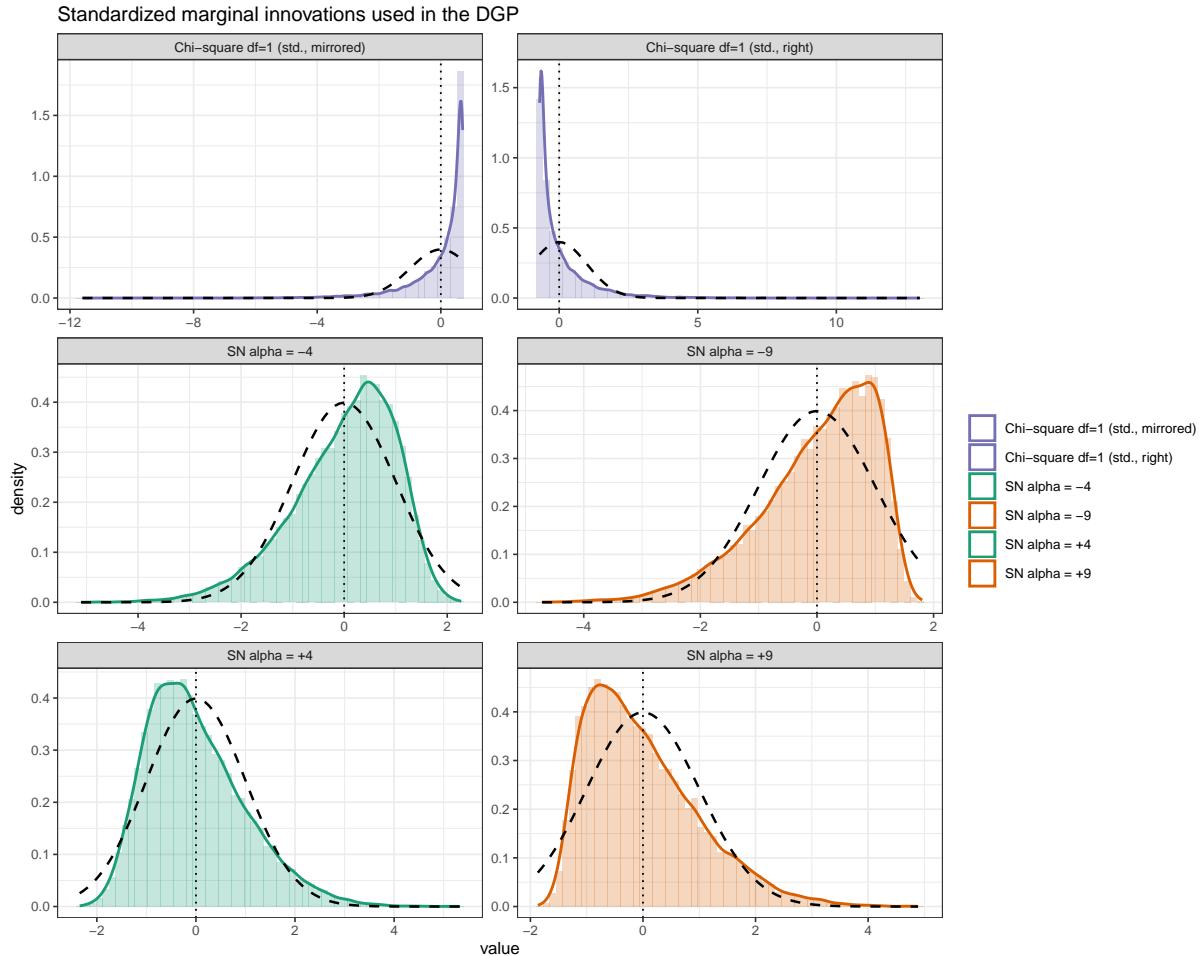
## 1.4 Visual Check: Standardized Marginal Innovations (DGP)

Standardized marginal innovations used in the DGP
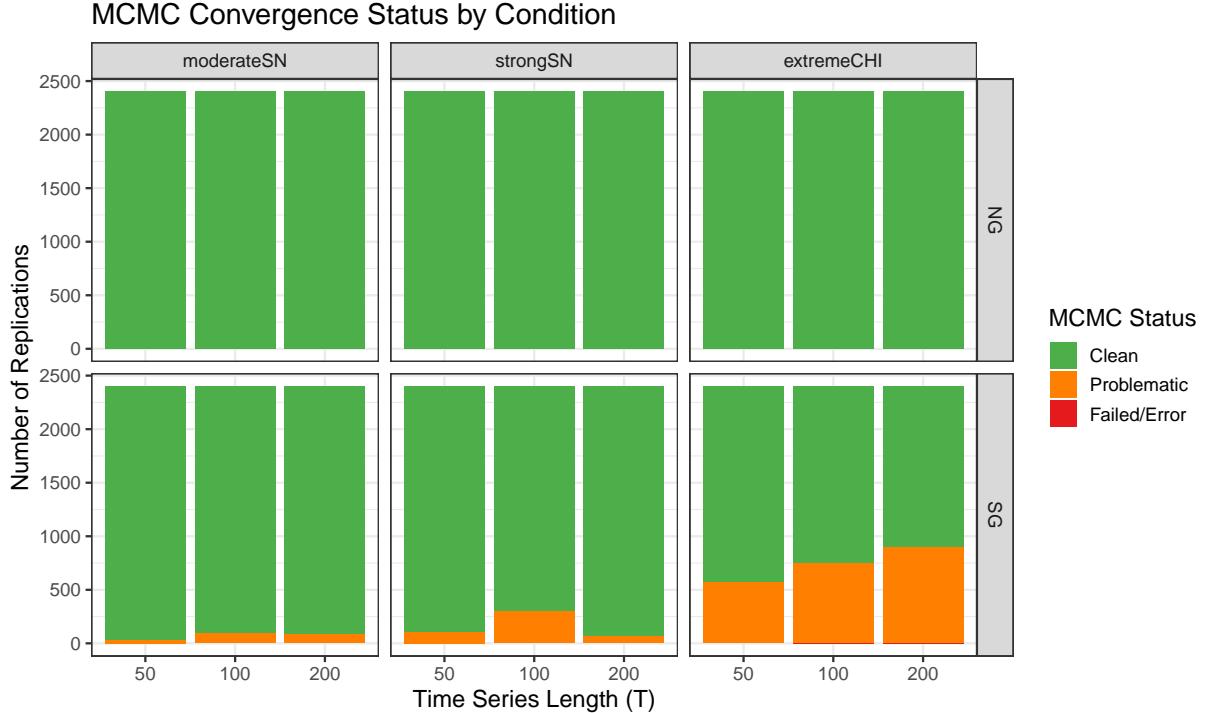
2. Data Loading and Preparation

2.1. MCMC Classification and Overview

We classify runs based on MCMC diagnostics and summarize the computational performance. A replication run was deemed as "Problematic" if it successfully completed sampling but exhibited either of the following conditions:

- **High R-hat:** The potential scale reduction factor (max_rhat) for any parameter was greater than 1.01.
- **Divergent Transitions:** The sampler reported one or more divergent transitions (n_div > 0) after warmup (if recorded).

MCMC Convergence Status by Condition

Interpretation: NG fits are uniformly Clean across the design. SG sampling degrades with increasing marginal skewness: most runs are Clean under moderateSN, Problematic rates are modest under strongSN, and under extremeCHI roughly one-quarter to one-third of runs exceed the $\hat{R}$ threshold (depending on T).

```
Divergent transition counts were not recorded for this render.
```

Interpretation: Divergence distributions are shown only when n_div is available. When divergence counts are missing, diagnostics reflect $\hat{R}$ rather than divergences.

2.2. Failure-Inclusive Summary (SG, extremeCHI)

We report the SG success rate under extremeCHI and the conditional coverage among successful fits. The failure-inclusive coverage treats failed fits as non-coverage by multiplying success rate by conditional coverage.

Table 3: SG extremeCHI: success rate and conditional coverage (failure-inclusive).

| param	conditional_coverage	success_rate	failure_inclusive_coverage
mu[1]	0.280	1	0.280
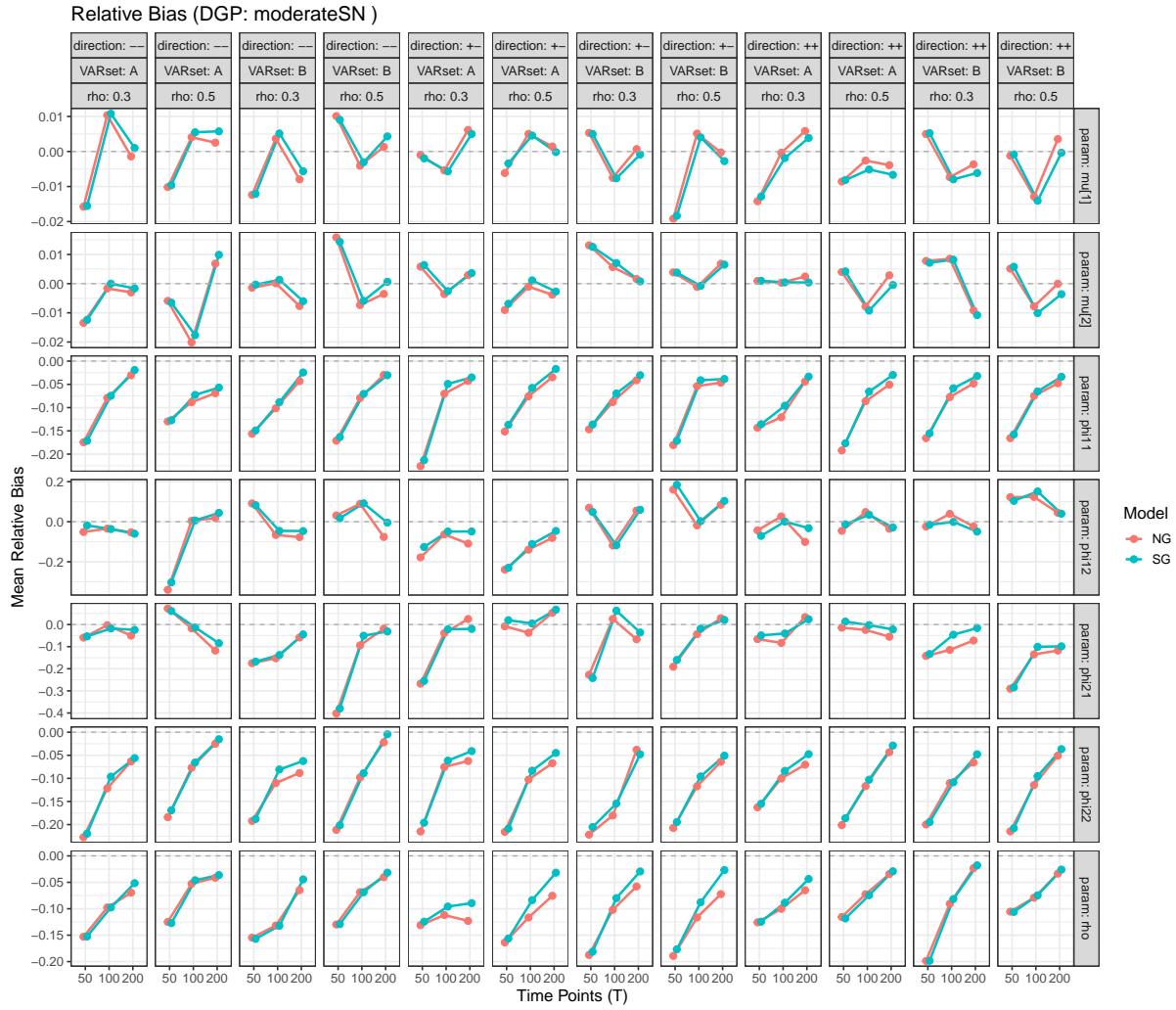
| param	conditional_coverage	success_rate	failure_inclusive_coverage
mu[2]	0.287	1	0.287
phi11	1.000	1	0.999
phi12	1.000	1	0.999
phi21	1.000	1	0.999
phi22	0.999	1	0.999
rho	0.573	1	0.573

4. Condition 1: Moderate Skewness (moderateSN)
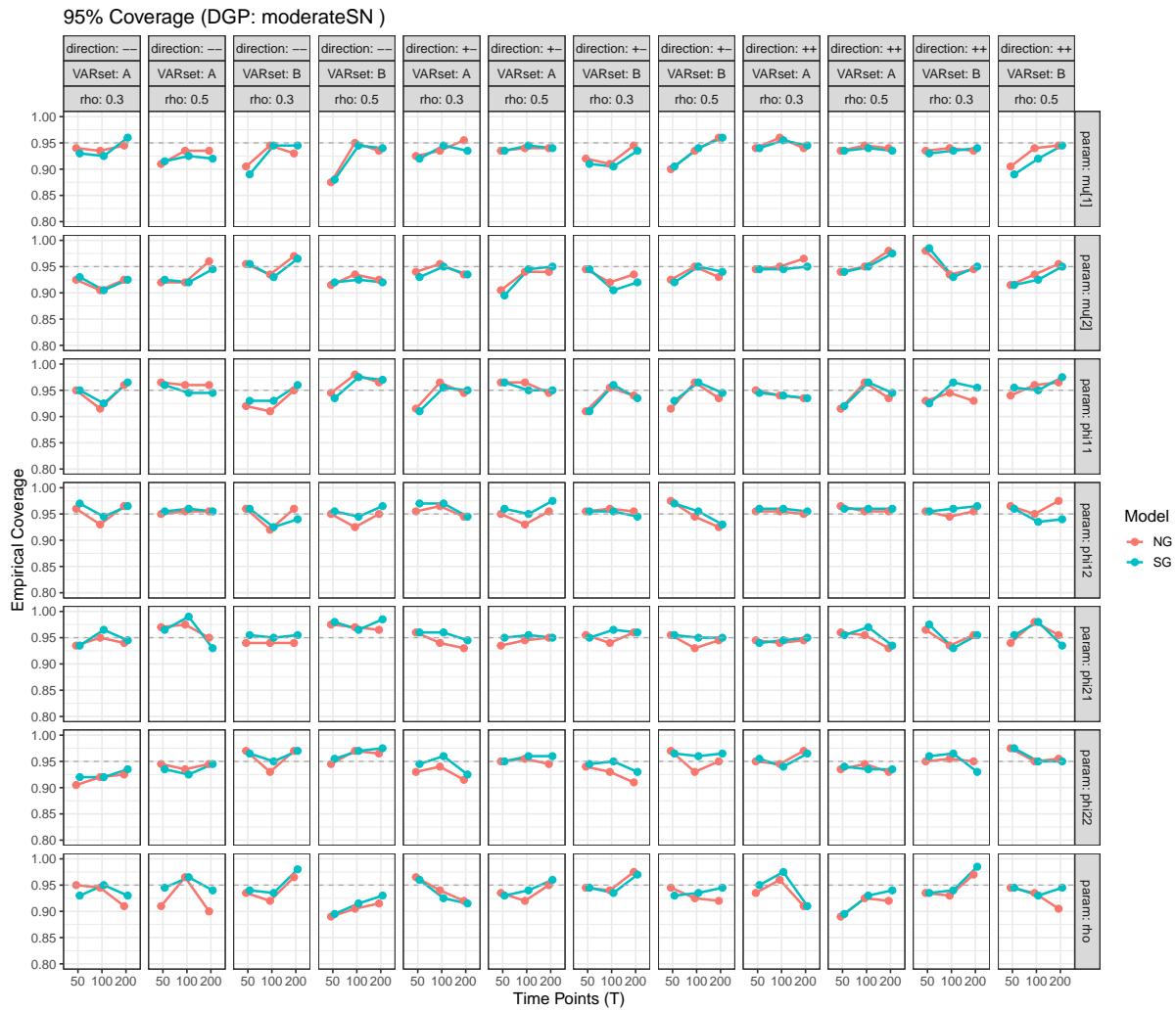
DGP: Skew-Normal innovations ($\alpha = 4$). NG model is misspecified; SG model is correctly specified.

```
plots_mod <- generate_plots_for_condition("moderateSN", cond_summary)
```
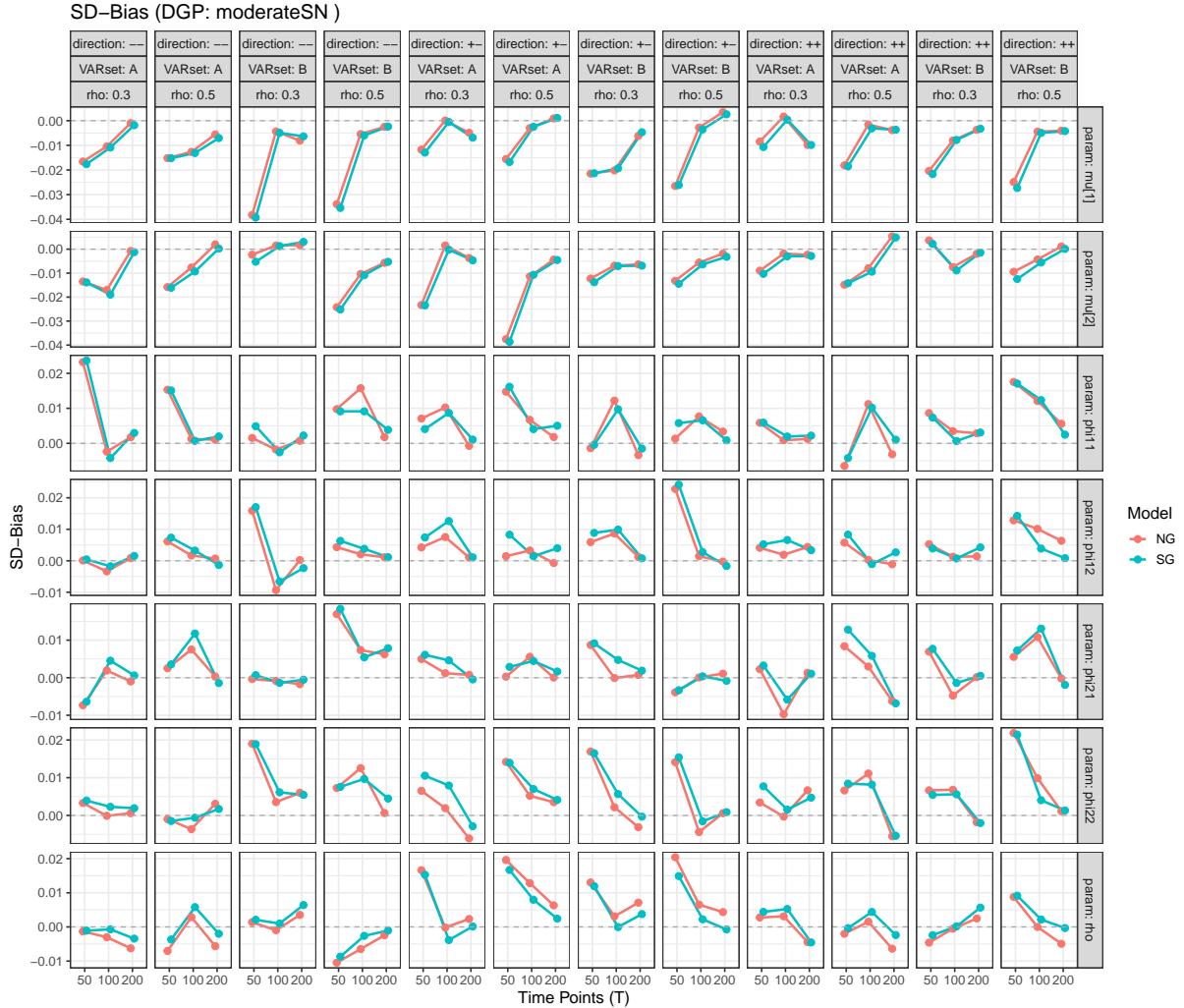
4.1. Relative Bias (moderateSN)

Relative Bias (DGP: moderateSN)

4.2. 95% Coverage (moderateSN)

95% Coverage (DGP: moderateSN)

4.3. SD-Bias (moderateSN)

SD−Bias (DGP: moderateSN)
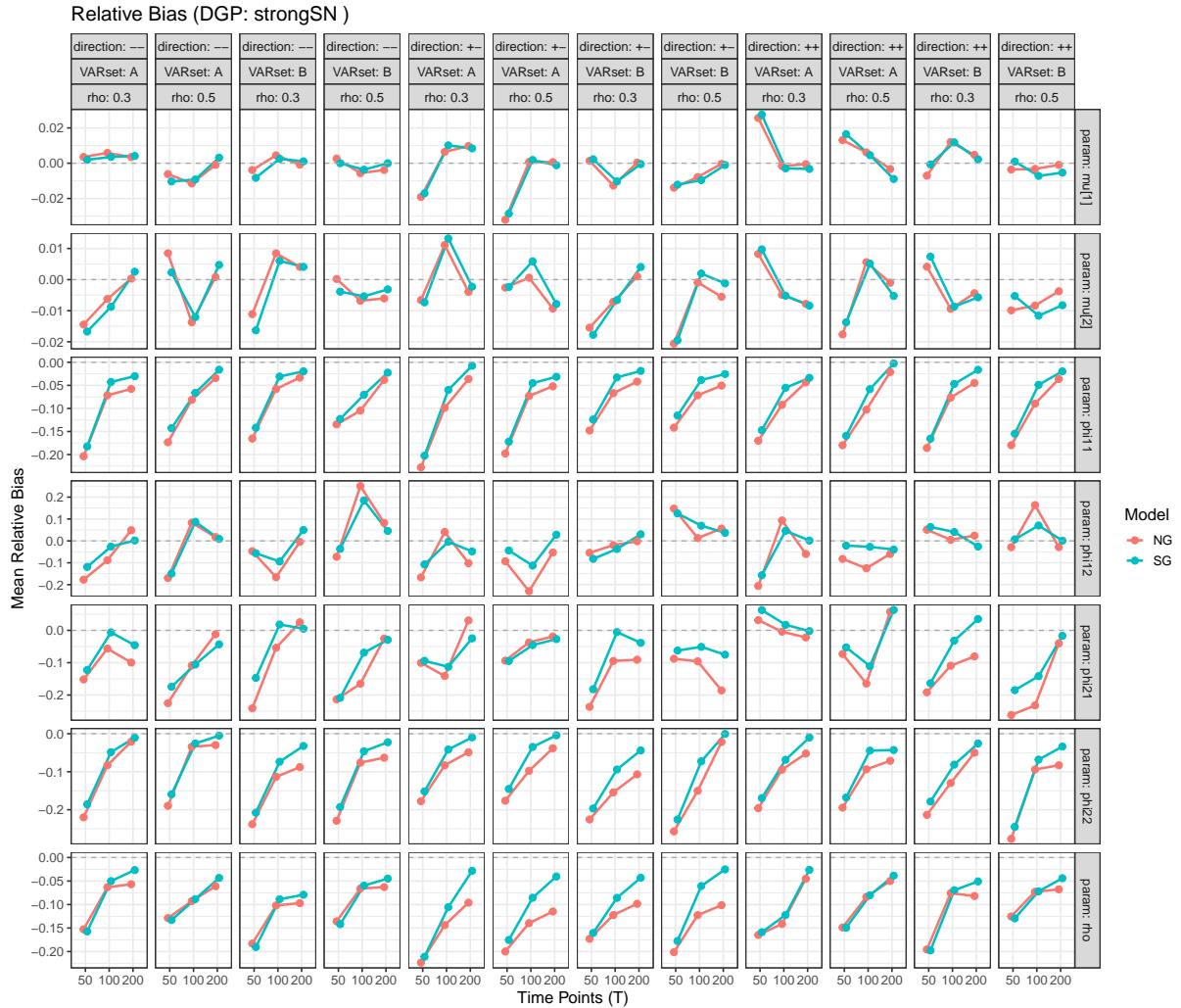
Interpretation (moderateSN): Under moderateSN, NG and SG have similar operating characteristics. Bias is small and coverage is near 0.95 for most parameters; modest finite-sample attenuation in some Φ and ρ panels is visible at $T = 50$ but largely disappears by $T \geq 100$. SD-bias is close to zero for both models, indicating reasonably calibrated posterior uncertainty. Under this level of skewness, SG's added complexity is not supported by material inferential gains.

5. Condition 2: Strong Skewness (strongSN)
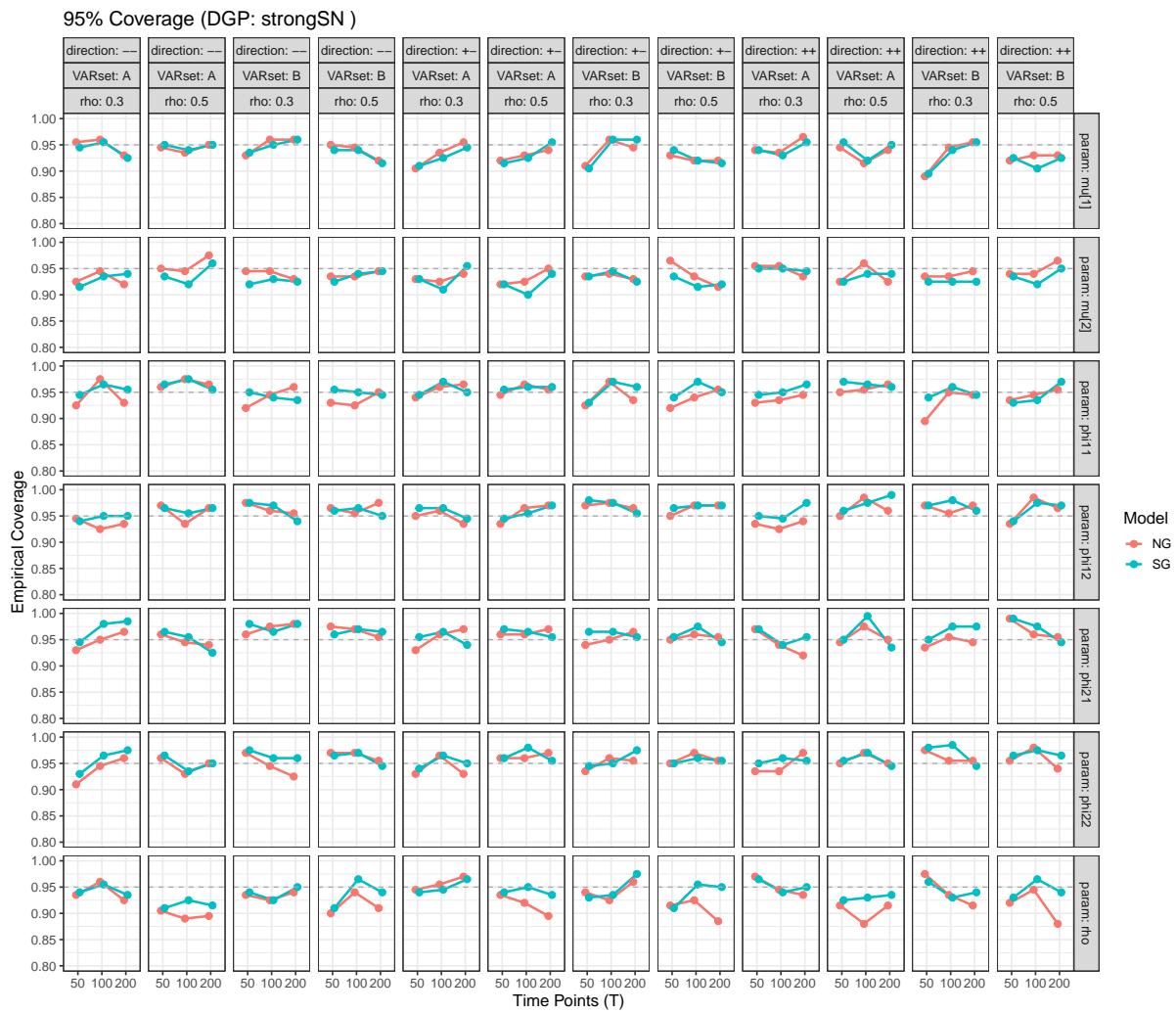
DGP: Skew-Normal innovations ($\alpha = 9$). The misspecification for the NG model is severe.

```
plots_strong <- generate_plots_for_condition("strongSN", cond_summary)
```
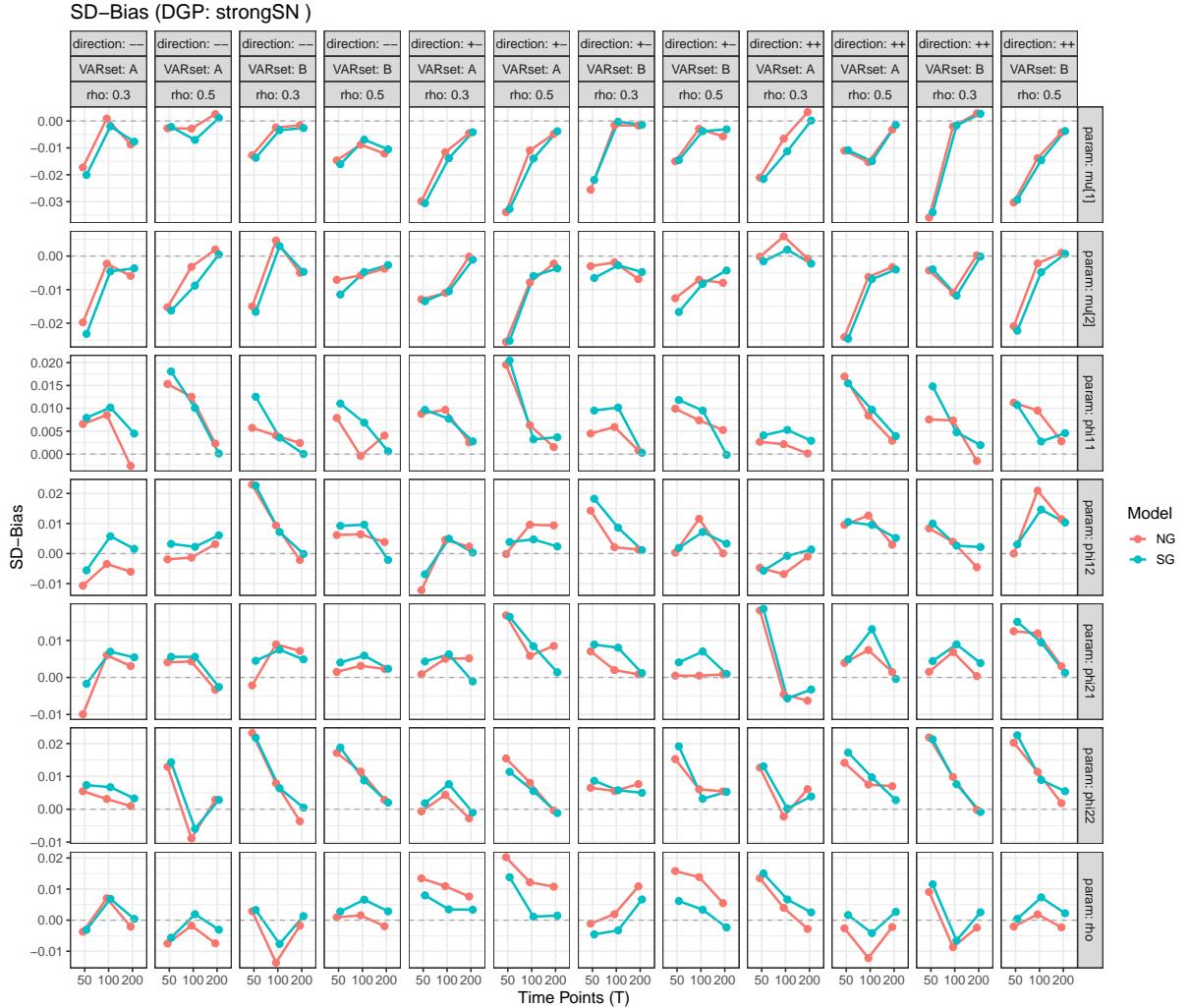
5.1. Relative Bias (strongSN)

Relative Bias (DGP: strongSN)

5.2. 95% Coverage (strongSN)

95% Coverage (DGP: strongSN)

5.3. SD-Bias (strongSN)

SD−Bias (DGP: strongSN)
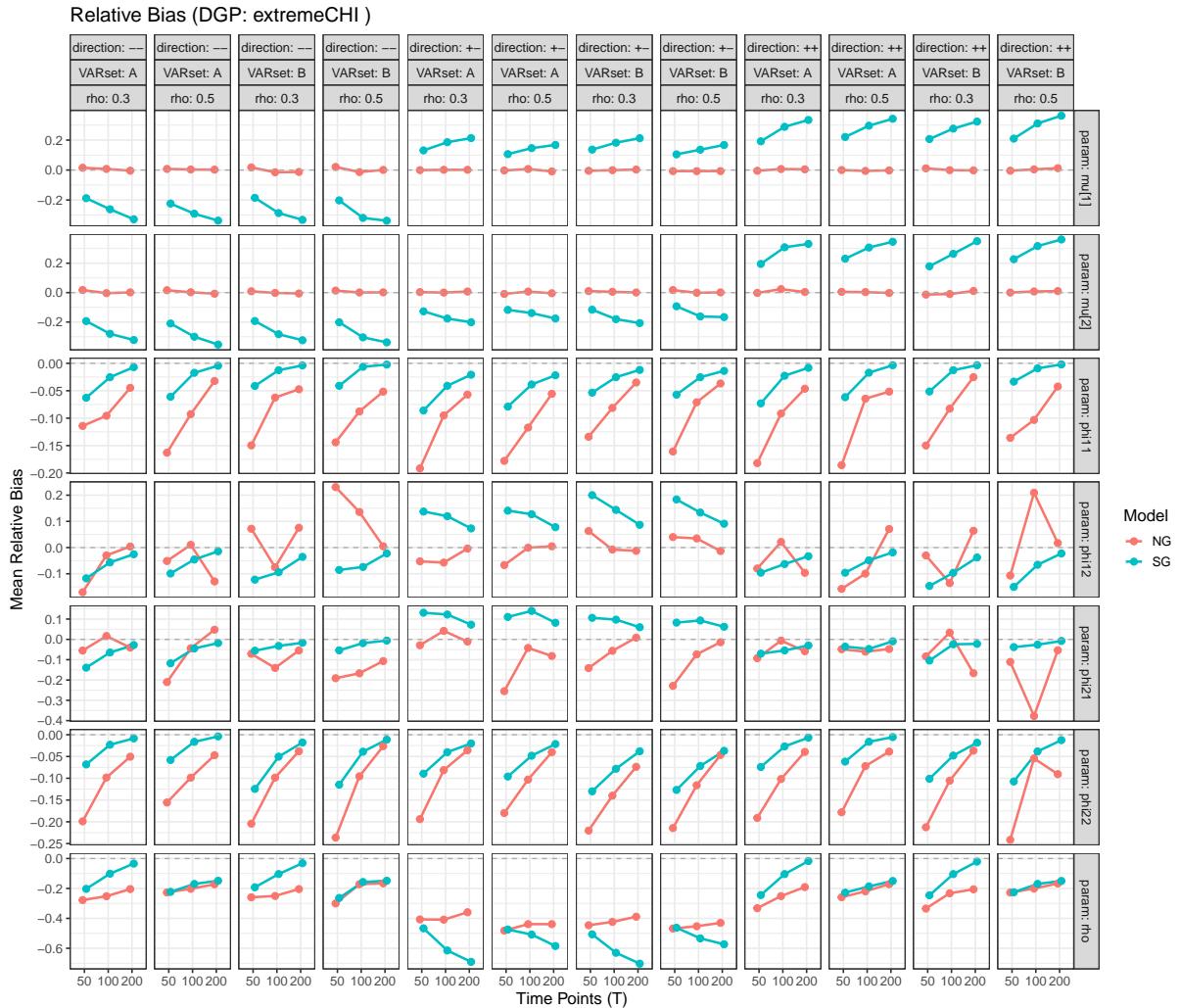
Interpretation (strongSN): Under strong skewness ($\alpha = \pm 9$), meaningful differences between the NG and SG models begin to emerge, though they remain modest. The SG model reduces bias for the VAR dynamics (Φ) on average (around 0.04 in relative bias, with occasional improvements up to ≈ 0.12). Coverage rates for both models remain near nominal for most parameters, with the SG model showing marginally better calibration. The copula parameter ρ exhibits downward bias under both models, though this is less severe than under extremeCHI. SD-Bias remains small for both models, indicating that posterior uncertainty quantification is reasonably well-calibrated even under strong skewness. The computational cost of the SG model (elevated $\hat{R}$ in some cells) may not be justified by the modest improvements in inference, particularly for shorter time series.

18

6. Condition 3: Extreme Skewness and Misspecification (extremeCHI)

DGP: standardized chi-squared ($df = 1$) innovations. Both models are marginally misspecified; SG allows skewness in the innovations but remains outside the χ_1^2 family.

```
plots_extreme <- generate_plots_for_condition("extremeCHI", cond_summary)
```

6.1. Relative Bias (extremeCHI)

Relative Bias (DGP: extremeCHI)

Interpretation: High Bias for SG Model in intercepts (μ_1, μ_2)

A salient feature of extremeCHI is the systematic bias in the SG intercepts (μ_1, μ_2), while the NG intercepts remain close to 0.

- **NG:** Although NG is severely misspecified in higher moments, the DGP innovations are standardized to have mean 0. In these simulations, NG's posterior for μ remains centered near 0, with misspecification primarily expressed through distorted estimates of Φ and, most prominently, ρ.

- **SG:** SG is also misspecified under extremeCHI because standardized χ_1^2 marginals lie far outside the skew-normal family (Table below). Under the centered skew-normal parameterization, the innovation distribution is constrained to have mean 0, but cannot reproduce the DGP tail behavior. Empirically, the posterior shifts μ away from 0 in some settings, suggesting that the intercept absorbs part of the marginal misspecification through joint estimation with (Φ, α, ω).

This shift in μ does not mitigate copula attenuation: ρ remains strongly biased under both models.

---

i Note

The extremeCHI condition uses standardized Chi-squared innovations with 1 degree of freedom $(\chi^2(1))$. The mismatch between this DGP and the model assumptions is strong:

Distribution	Theoretical Skewness	Theoretical Excess Kurtosis
True DGP $(\chi^2(1))$	≈ 2.83	12
NG Model (Normal)	0	0
SG Model (Skew-Normal Max)	≈ 0.995	≈ 0.869

---

i Skewness and kurtosis formulas

- χ_ν^2: $\mathbb{E} = \nu$, $\mathrm{Var} = 2\nu$, $\gamma_1 = \sqrt{8/\nu}$, $\gamma_2 = 12/\nu$ (excess). For $\nu = 1$: $\gamma_1 \approx 2.828$, $\gamma_2 = 12$.

- Skew-normal $\mathrm{SN}(\xi, \omega, \alpha)$: with $\delta = \alpha/\sqrt{1+\alpha^2}$,

$$\mu = \xi + \omega\delta\sqrt{\frac{2}{\pi}}, \quad \sigma^2 = \omega^2\left(1 - \frac{2\delta^2}{\pi}\right),$$

$$\gamma_1 = \frac{(4-\pi)}{2}\frac{(\delta\sqrt{2/\pi})^3}{(1-2\delta^2/\pi)^{3/2}}, \quad \gamma_2 = 2(\pi-3)\frac{(\delta\sqrt{2/\pi})^4}{(1-2\delta^2/\pi)^2}.$$

As $|\alpha| \to \infty$ (i.e., $|\delta| \to 1$), the maximum skewness is $\gamma_1 \approx 0.995$ and the maximum excess kurtosis is $\gamma_2 \approx 0.869$. This is the basis for the entries in the table contrasting χ_1^2 vs. Normal vs. Skew-Normal.

Mechanism: PIT distortion under marginal misspecification. The PIT maps each margin Y to $U = F_{\text{assumed}}(Y)$. If $F_{\text{assumed}} = F_{\text{true}}$, then $U \sim \text{Uniform}(0,1)$. Under misspecification:

- **Tail compression:** If the true data have heavier right tails (e.g., χ_1^2) than the assumed SN/Normal, then very large Y values do not land near 1 after transformation; they are pulled back toward the center (e.g., $U \approx 0.8$ instead of 0.98). Left tails are similarly distorted under mirroring.

- **Rank distortion:** The PIT is a ranking device. By compressing real extremes toward the middle, co-extreme events (Y_1, Y_2) that truly move together in the tails no longer co-locate in the corners of $[0,1]^2$; instead they fall into the interior.
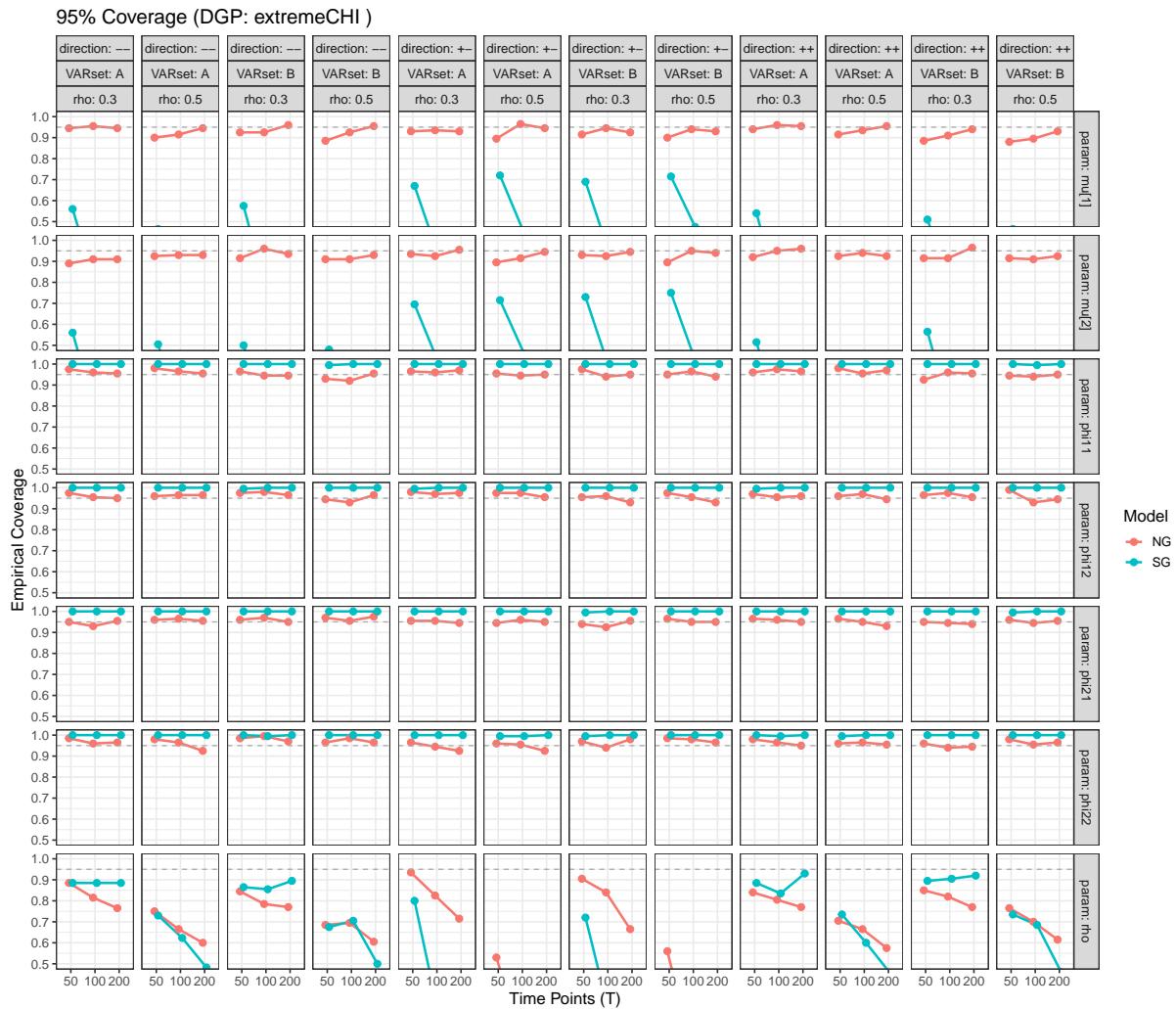
For Gaussian copulas, dependence is most visible in the corners. When misspecified marginals push mass away from corners, the copula "sees" less tail co-movement even if it exists in the data, and any fitted ρ is forced downward. This pattern is reflected in the ρ relative-bias panels.

Mechanism: attenuation of ρ. The joint likelihood factorizes as
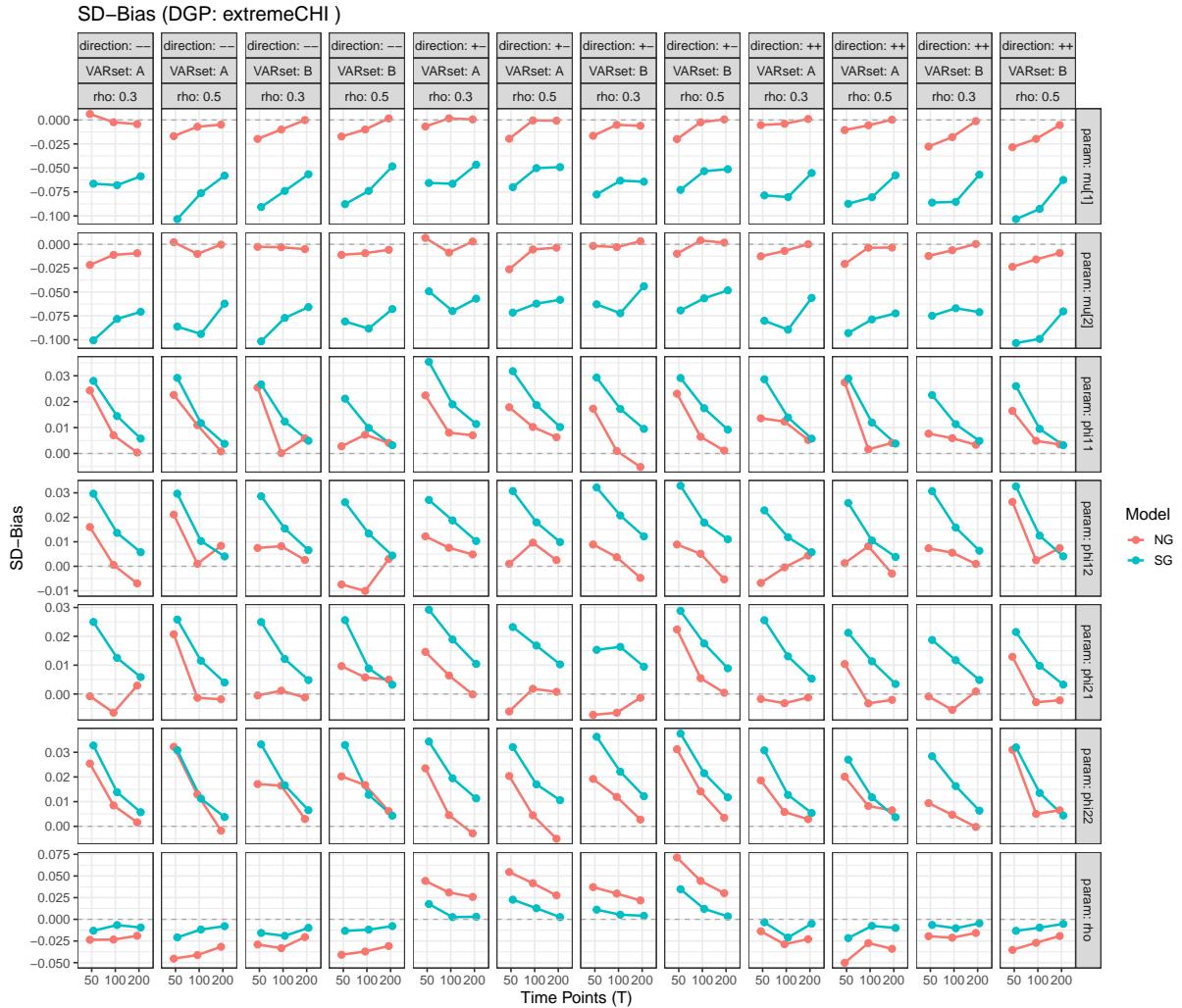
$$\prod_t c(U_{1,t}, U_{2,t}; \rho) f_1(y_{1,t}) f_2(y_{2,t}),$$

so only the copula term $c(\cdot; \rho)$ can adjust dependence. When the PIT pushes tail pairs $(U_{1,t}, U_{2,t})$ toward the center ($\approx 0.5, 0.5$), the Gaussian copula density becomes less sensitive to ρ (the score in ρ flattens). To avoid penalizing improbable tail corners that the distorted (U_1, U_2) no longer occupy, the MLE/posterior moves ρ down toward 0. Hence the large negative relative bias and sub-nominal coverage for ρ in extremeCHI, regardless of "Clean" or "Problematic" MCMC status.

6.2. 95% Coverage (extremeCHI)

95% Coverage (DGP: extremeCHI)

6.3. SD-Bias (extremeCHI)

SD–Bias (DGP: extremeCHI)

Interpretation (extremeCHI): Severe marginal misspecification produces parameter-specific failures.

- **NG:** μ remains approximately unbiased, but Φ and especially ρ exhibit substantial bias. Coverage for Φ is close to nominal on average but drops in the worst cells; some replications show posterior means for ρ crossing zero (relative bias < -1).

- **SG:** Φ is comparatively robust (bias smaller), but coverage is conservative (often > 0.99 for $T \geq 100$). μ is biased and ρ remains severely attenuated.

The shared failure for ρ is consistent with PIT distortion under extreme mismatch between the assumed marginal CDF (Normal or skew-normal) and the standardized χ_1^2 DGP. SD-bias

for the dynamic parameters is often close to zero, indicating that posterior uncertainty for Φ can remain reasonably calibrated even when ρ (and, for SG, μ) is biased.

6.4. PIT Distortion Evidence (NG) and ρ Attenuation

To link PIT distortion to ρ attenuation, we compute a simple PIT diagnostic under the NG marginal assumption. For a random subset of simulated series per skew regime, we apply the standard Normal CDF to the true residuals and report the KS statistic against Uniform(0, 1). We then compare this PIT distortion to the average ρ relative bias for NG.

PIT Distortion (NG) vs. rho Attenuation
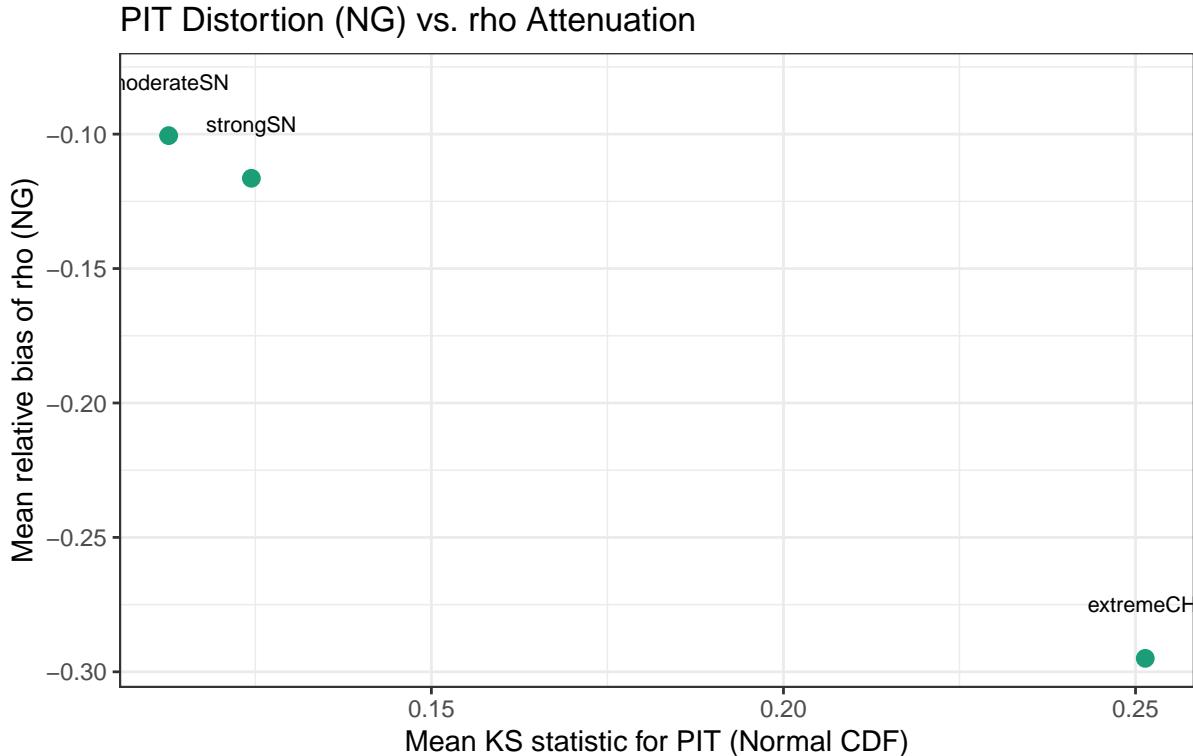
Table 5: PIT distortion and rho attenuation by skew regime (NG).
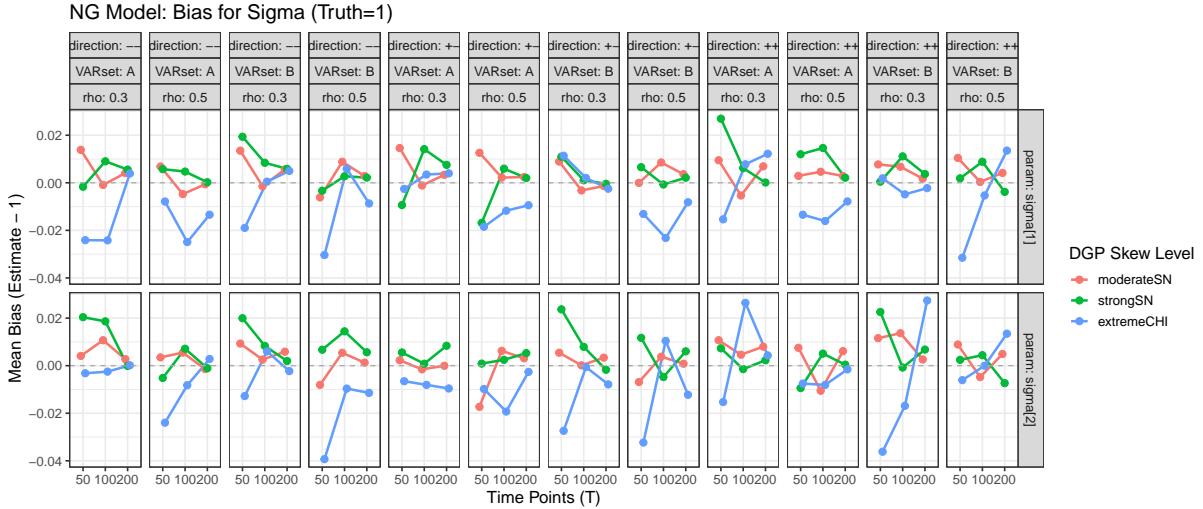
skew_level	pit_ks	rho_bias
extremeCHI	0.251	-0.295
moderateSN	0.113	-0.101
strongSN	0.124	-0.116

7. Marginal Parameters

We examine the parameters governing the marginal distributions to understand the mechanisms driving the results.

7.1. NG Scale Parameter Behavior (σ)

The innovations are standardized to unit variance. Under correct specification, the NG model should recover $\sigma = 1$.

NG Model: Bias for Sigma (Truth=1)

Interpretation: The σ estimates under NG show small but systematic deviations from the true value of 1 (bias $\approx \pm 0.025$). These deviations represent a byproduct of marginal misspecification: when the NG model encounters skewed or heavy-tailed innovations, it cannot capture the shape mismatch and instead makes minor adjustments to the scale parameter.
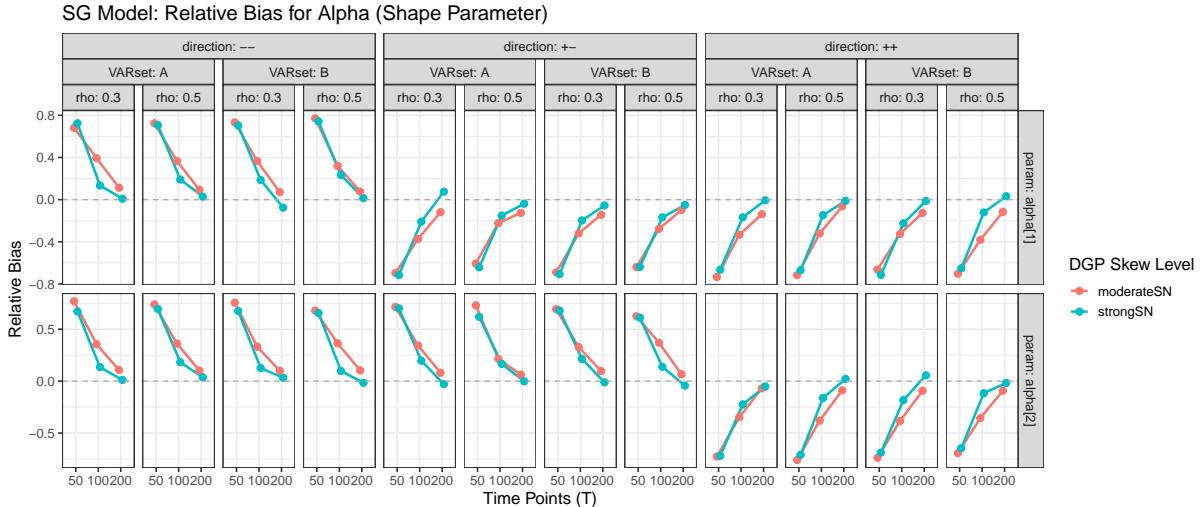
Importantly, σ **bias is not the mechanism driving ρ attenuation.** The two phenomena arise from the same root cause (marginal misspecification) but through different pathways:

- σ **bias:** The NG model slightly adjusts scale to improve marginal likelihood under non-Gaussian data. This is a local accommodation with limited impact on other parameters.

- ρ **attenuation:** Arises from PIT distortion—regardless of how well the marginal density fits, if the marginal CDF misrepresents tail probabilities, the copula receives distorted inputs. Even with σ perfectly estimated, the PIT would still compress tail observations toward the center of $[0,1]^2$, attenuating perceived dependence.

The small magnitude of σ bias ($\approx 2.5\%$) compared to the large ρ bias (often $> 100\%$) confirms that these are parallel consequences of misspecification rather than a causal chain where σ error propagates to ρ.

7.2. SG Shape Parameter Recovery (α)

We examine how well the SG model recovers the true shape parameter α (applicable for SN conditions).

SG Model: Relative Bias for Alpha (Shape Parameter)

Interpretation (shape recovery). Under the correctly specified SN DGPs, posterior means for α are biased toward 0 at $T = 50$, with substantially reduced bias at $T \geq 100$. Despite imperfect recovery of α in short series, allowing skewness in the marginals appears sufficient to stabilize estimation of the VAR dynamics relative to NG.

> **!** Caveat: large $|\alpha|$ is weakly identified at short T
>
> For $|\alpha| \in \{4, 9\}$, the transformed shape parameter $\delta = \alpha/\sqrt{1 + \alpha^2}$ is close to the boundary ± 1. In this regime, α and ω are strongly confounded and the likelihood contains limited information about the exact magnitude of α at $T = 50$. With regularizing priors, posterior estimates can therefore shrink toward 0 even under correct specification. The observed downward bias in $|\alpha|$ at $T = 50$ should be interpreted primarily as finite-sample regularization, not as a failure to recover the direction of skewness.
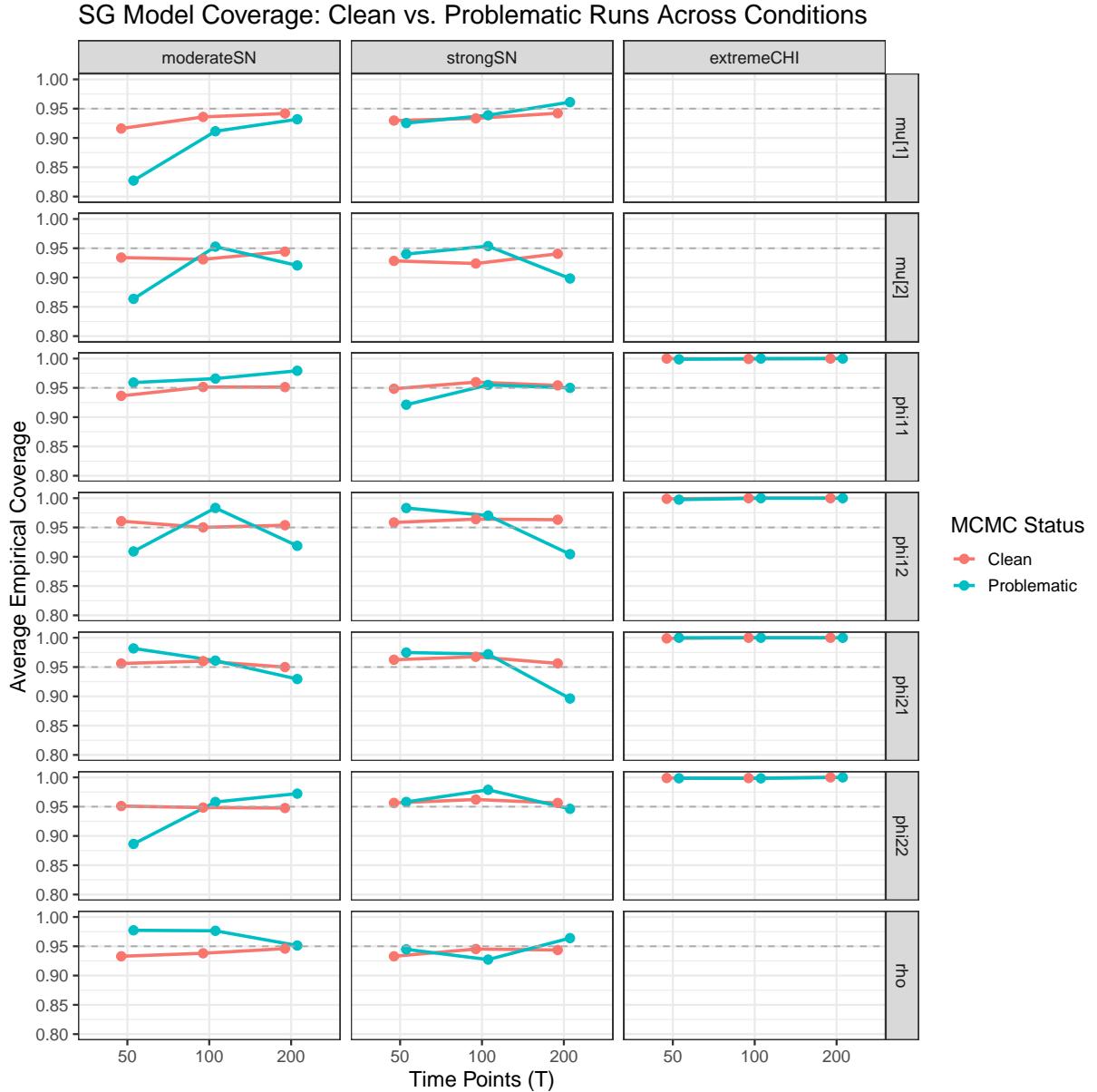
Mechanistic interpretation. Even when $|\alpha|$ is underestimated at $T = 50$, the sign is typically correct, which partially aligns the PIT with the data. This partial correction is consistent with the small bias and near-nominal coverage for Φ under SG in the SN DGPs, and with the improvement in Φ performance as T increases and α becomes better identified.

8. Impact of MCMC Diagnostics

The SG model frequently encountered "Problematic" MCMC runs. We investigate if the statistical performance differs between "Clean" and "Problematic" runs for the SG model.

8.1. Coverage Split by MCMC Status (SG Model)

We visualize the coverage across all three conditions, comparing Clean vs. Problematic runs.

SG Model Coverage: Clean vs. Problematic Runs Across Conditions

Interpretation: Impact of MCMC Status Across this design, SG coverage for Φ is similar in the Clean and Problematic subsets; even under extremeCHI, coverage for the dynamic parameters remains near nominal, whereas coverage for μ deteriorates (Section 6). This indicates that—at least for the functionals assessed here—the diagnostic rule used to define Problematic status is not, on average, strongly associated with degraded Φ coverage.

> 🔥 Caveats on interpreting "Clean" vs. "Problematic" comparisons
>
> 1. **Problematic is a composite label** ($\hat{R} > 1.01$ and/or $n_{\text{div}} > 0$). Pooling these sources of pathology can dilute relationships between diagnostics and inferential quality.
>
> 2. **Coverage is computed conditional on runs that completed sampling**; it does not account for selection induced by failed fits.
>
> 3. **The summaries average over ρ, VAR set, and skewness direction.** Localized failures in specific design cells may be obscured by aggregation.
>
> 4. **Divergences can bias some posterior functionals** even when marginal coverage for a subset of parameters appears adequate; when available, they should continue to be treated as a warning sign requiring model reparameterization or stricter sampling settings.

8.2. Relationship between Bias and Divergences

We examine if runs with more divergences exhibit higher bias at the replication level. We focus on the SG model under strongSN conditions (only when divergence counts are available).

```
Divergence counts were not recorded for this render; skipping this diagnostic.
```

Interpretation: Bias vs. Divergences When divergence counts are available, there does not appear to be a strong correlation between the number of divergences and the absolute bias for the core parameters. High bias occurs in runs with few divergences, and low bias occurs in runs with many divergences. This reinforces the conclusion that MCMC diagnostics, while important indicators of computational issues, are not reliable predictors of statistical accuracy in this context.

10. Details

This section provides technical details on the implementation of the simulation study, including prior specifications, MCMC settings, and mathematical derivations.

10.1. Prior Specifications

Both models use weakly informative priors designed to regularize estimates while allowing the data to dominate inference. The priors are summarized in the following tables.

Table: Prior Specifications for the Normal-Gaussian (NG) Model

Parameter	Prior	Support	Rationale
μ_1, μ_2	Normal$(0, 1)$	$\mathbb{R}$	Weakly informative; centered at true value (0)
$\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}$	Normal$(0, 0.5)$	$(-1, 1)$	Regularizes toward stationarity; truncated by bounds
σ_1, σ_2	Half-Normal$(0, 1)$	$(0, \infty)$	Weakly informative scale prior; mode near 0, mass around 1
ρ	Normal$(0, 0.5)$	$(-1, 1)$	Regularizes toward independence; truncated by bounds

Table: Prior Specifications for the Skew-Normal-Gaussian (SG) Model

Parameter	Prior	Support	Rationale
μ_1, μ_2	Normal$(0, 1)$	$\mathbb{R}$	Weakly informative; centered at true value (0)
$\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}$	Normal$(0, 0.5)$	$(-1, 1)$	Regularizes toward stationarity; truncated by bounds
ω_1, ω_2	Half-Normal$(0, 1)$	$(0, \infty)$	Scale parameter; concentrates mass near 1
δ_1, δ_2	Normal$(0, 0.5)$	$(-1, 1)$	Regularizes toward symmetry ($\delta = 0$); truncated by bounds
ρ	Normal$(0, 0.5)$	$(-1, 1)$	Regularizes toward independence; truncated by bounds

> **ℹ Half-Normal Priors**
>
> For parameters constrained to be positive (e.g., σ, ω), Stan automatically truncates the normal distribution at 0, yielding a half-normal prior. A Half-Normal$(0, 1)$ prior has mode 0, mean $\sqrt{2/\pi} \approx 0.80$, and standard deviation $\sqrt{1 - 2/\pi} \approx 0.60$.

10.2. MCMC Settings

All models were fitted using the No-U-Turn Sampler (NUTS) implemented in Stan via the `rstan` package. The following settings were used:

Table: MCMC Sampling Configuration

Setting	Value	Description
Chains	4	Number of independent Markov chains
Total iterations	4,000	Iterations per chain (including warmup)
Warmup iterations	2,000	Discarded adaptation period
Post-warmup draws	2,000	Retained samples per chain
`adapt_delta`	0.95	Target acceptance probability (higher = smaller step size)
`max_treedepth`	15	Maximum tree depth for NUTS
Parallelization	Outer loop	Replications parallelized; chains run sequentially

> **ℹ Convergence Diagnostics**
>
> A replication was classified as **Problematic** if either:
>
> - Maximum $\hat{R} > 1.01$ across all parameters, or
> - Number of post-warmup divergent transitions $n_{\text{div}} > 0$ (when recorded)
>
> The elevated `adapt_delta = 0.95` (default is 0.80) was chosen to reduce divergent transitions in the SG model, which exhibits more challenging posterior geometry due to the skew-normal parameterization.

10.3. Gaussian Copula Log-Density

The Gaussian copula density for uniform marginals $(u, v) \in (0, 1)^2$ with correlation parameter $\rho \in (-1, 1)$ is implemented as follows.

Let $z_1 = \Phi^{-1}(u)$ and $z_2 = \Phi^{-1}(v)$ denote the standard normal quantile transforms. The copula log-density is:

$$\log c(u, v; \rho) = -\frac{1}{2}\log(1 - \rho^2) - \frac{1}{2(1 - \rho^2)}\left(z_1^2 - 2\rho z_1 z_2 + z_2^2\right) + \frac{1}{2}\left(z_1^2 + z_2^2\right)$$

This expression is derived from the bivariate normal density minus the product of marginal standard normal densities:

$$c(u, v; \rho) = \frac{\phi_2(z_1, z_2; \rho)}{\phi(z_1)\phi(z_2)}$$

where $\phi_2(\cdot, \cdot; \rho)$ is the bivariate standard normal density with correlation ρ, and $\phi(\cdot)$ is the univariate standard normal density.

---

i Numerical Stability: Boundary Clamping

The quantile function $\Phi^{-1}(u)$ diverges as $u \to 0$ or $u \to 1$. To prevent numerical overflow, the implementation applies boundary clamping:

$$u_{\text{clamped}} = \max(\varepsilon, \min(1 - \varepsilon, u)), \quad \varepsilon = 10^{-9}$$

This affects only observations beyond the 0.9999999th percentile and has negligible impact on inference. The Stan implementation:

```
real gaussian_copula_ld(real u, real v, real rho) {
  real eps = 1e-9;
  real uu  = fmax(eps, fmin(1 - eps, u));
  real vv  = fmax(eps, fmin(1 - eps, v));
  real z1  = inv_Phi(uu);
  real z2  = inv_Phi(vv);
  real rho2 = square(rho);

  return -0.5 * log1m(rho2)
         - 0.5 / (1 - rho2) * (square(z1) - 2 * rho * z1 * z2 + square(z2))
         + 0.5 * (square(z1) + square(z2));
}
```

10.4. Chi-Squared Mirroring for Left-Skewed Marginals

The `extremeCHI` condition generates innovations from standardized χ_1^2 distributions. For left-skewed marginals (directions `--` and `+-`), the distribution must be mirrored while preserving

the copula correlation structure.

Naive approach (incorrect): Given $U \sim \text{Uniform}(0,1)$ from the copula, compute $-F_{\chi^2}^{-1}(U)$ to obtain left-skewed draws.

Problem: This approach flips the sign of the copula correlation. If $\rho > 0$, large U_1 tends to co-occur with large U_2. Under the naive mirroring, large U_1 maps to large negative ϵ_1, but large U_2 maps to large positive ϵ_2, reversing the dependence.

Correct approach: For the mirrored (left-skewed) marginal, use the quantile function of $-X$:

$$Q_{-X}(u) = -Q_X(1 - u)$$

This transforms the uniform *before* applying the quantile function, preserving the copula structure:

```
# Correct implementation in simulate_data.R
function(pu) {
  # Adjust uniform FIRST if mirroring
  pu_adj <- if (mirror) 1 - pu else pu
  x_raw <- qchisq(pu_adj, df)
  x_std <- (x_raw - mean_chi) / sd_chi
  # Then negate for left-skewed distribution
  if (mirror) -x_std else x_std
}
```

With this implementation, the copula parameter ρ is preserved across directions, so bias and coverage summaries use the input ρ directly.

> **!** Verification via Kendall's Tau
>
> The simulation pipeline verifies copula correlation preservation by computing empirical Kendall's τ on generated residuals:
>
> $$\tau_{\text{expected}} = \frac{2}{\pi} \arcsin(\rho)$$
>
> For a Gaussian copula with $\rho = 0.5$, this yields $\tau \approx 0.333$. The pipeline flags conditions where $|\tau_{\text{empirical}} - \tau_{\text{expected}}|$ exceeds $3 \times \text{SE}(\tau)$, where $\text{SE}(\tau) \approx \sqrt{2(2n+5)/(9n(n-1))}$ for sample size n.

10.5. Centered Parameterization for Skew-Normal Marginals

The SG model uses a **centered parameterization** (CP) that enforces zero-mean innovations by construction. This section documents the transformation from CP parameters (δ, ω) to the **direct parameterization** (DP) (ξ, ω, α) required by Stan's `skew_normal` functions.

Parameterization Relationship

The skew-normal distribution $\text{SN}(\xi, \omega, \alpha)$ has:

- Location: $\xi \in \mathbb{R}$
- Scale: $\omega > 0$
- Shape: $\alpha \in \mathbb{R}$

The derived parameter δ relates to α via:

$$\delta = \frac{\alpha}{\sqrt{1 + \alpha^2}} \in (-1, 1)$$

with inverse:

$$\alpha = \frac{\delta}{\sqrt{1 - \delta^2}}$$

Centering Constraint

The mean of $\text{SN}(\xi, \omega, \alpha)$ is:

$$\mathbb{E}[X] = \xi + \omega\delta\sqrt{\frac{2}{\pi}}$$

To enforce $\mathbb{E}[\epsilon] = 0$, we set:

$$\xi = -\omega\delta\sqrt{\frac{2}{\pi}}$$

Stan Implementation

The SG model estimates (δ, ω) directly and computes (ξ, α) as transformed parameters:

```
transformed parameters {
  vector[2] alpha;
  vector[2] xi;

  // Transform delta -> alpha
  alpha = delta ./ sqrt(1 - square(delta));

  // Compute centering location
  xi = -omega .* (delta * sqrt(2.0 / pi()));
}
```

> **i Why Estimate δ Instead of α?**
>
> The parameter $\delta \in (-1, 1)$ is bounded, which facilitates prior specification and avoids the unbounded support of $\alpha \in \mathbb{R}$. As $|\alpha| \to \infty$, $|\delta| \to 1$, and small changes in δ near the boundary correspond to large changes in α. The prior $\delta \sim \text{Normal}(0, 0.5)$ (truncated to $(-1, 1)$) regularizes toward symmetry while permitting substantial skewness.

Variance Under the Centered Parameterization

The variance of $\text{SN}(\xi, \omega, \alpha)$ is:

$$\text{Var}(X) = \omega^2 \left(1 - \frac{2\delta^2}{\pi} \right)$$

In the DGP, we fix $\text{Var}(\epsilon) = 1$ by setting $\omega = (1 - 2\delta^2/\pi)^{-1/2}$. In the SG model, ω is estimated freely (see Section 1.1, Note on "SG Model Parameterization: Variance Not Fixed").

10.6. Reproducibility Strategy

The simulation study employs deterministic seeding to ensure full reproducibility regardless of execution order, parallelization, or resumption after interruption.

Data Generation Seeds

Each simulated dataset receives a unique seed derived from its position in the design:

$$\text{seed}_{\text{data}} = \text{seed\_base} + \text{condition\_id} \times 10000 + \text{rep\_id}$$

with seed_base $= 2{,}025{,}000$. This ensures:

- Different conditions and replications use non-overlapping seed ranges
- Regenerating a specific dataset always produces identical data
- Parallel execution order does not affect results

MCMC Initialization Seeds

Initial values for MCMC chains are generated deterministically:

$$\text{seed}_{\text{init}} = \text{seed\_base}_R + \text{model\_offset} + \text{condition\_id} \times 10000 + \text{rep\_id} \times 10 + \text{chain}$$

where:

- seed_base$_R = 1{,}000{,}000$
- model_offset $= 100$ for SG, 200 for NG
- chain $\in \{1, 2, 3, 4\}$

Stan Sampler Seeds

The Stan sampler receives seeds:

$$\text{seed}_{\text{Stan}} = \text{seed\_base}_{\text{Stan}} + \text{condition\_id} \times 1000 + \text{rep\_id}$$

with seed_base$_{\text{Stan}} = 5{,}000{,}000$.

> **i** Initialization Sidecar Files
>
> For each fitted model, the pipeline saves a "sidecar" file containing:
>
> - All random seeds used (data, init, Stan)
> - Exact initial values for each chain
> - Timestamp and metadata
>
> This enables exact reproduction of any individual fit and facilitates debugging of problematic runs.

Resume Capability

The pipeline supports resumption via environment variables:

```
START_COND <- as.integer(Sys.getenv("START_COND", "1"))
START_REP  <- as.integer(Sys.getenv("START_REP", "1"))
```

Because seeds are deterministic, resuming from condition 50, replication 100 produces identical results to running the full pipeline—the skipped conditions do not consume random numbers that would affect later computations.