# Bayesian unification of GLM, SEM, and DSEM: from Wright to PyMC

**Statistical modeling has converged.** What once appeared as distinct traditions—GLM/GLMM from biostatistics, SEM from psychometrics, and DSEM for intensive longitudinal data—are now understood as special cases of Bayesian hierarchical models. This unification, crystallized in Skrondal and Rabe-Hesketh's 2004 GLLAMM framework, (Lcbc-uio) reveals that **random effects ARE latent variables**: (Routledge) (Amazon) both are unobserved quantities drawn from population distributions and estimated through identical computational machinery. For practitioners building causal inference systems combining LLM-based DAG specification with Bayesian estimation, this convergence provides both theoretical foundation and practical implementation paths through PyMC's probabilistic programming framework.

The key insight is a shift from asking "which technique should I use?" to "what is the generative process that created my data?" Once you specify your beliefs about data generation—distributions, dependencies, hierarchical structure—the Bayesian framework handles estimation uniformly regardless of whether the model resembles a traditional mixed-effects regression, factor analysis, or time-series model.

## A century of convergence from Wright's guinea pigs to modern DSEM

The unification story begins with **Sewall Wright**'s invention of path analysis while still a graduate student around 1918. (nih) His 1920 paper on guinea pig coat patterns published the first path diagrams, and his 1921 "Correlation and Causation" formally described the method of path coefficients (nih) with "tracing rules" for decomposing correlations. Wright's approach was not merely regression—he used latent variables ("hypothetical factors") from the start. His 1934 paper in *Annals of Mathematical Statistics* demonstrated the method's full generality.

Wright's causal graphical approach faced resistance from **Karl Pearson**'s positivist rejection of causation (ResearchGate) and **Ronald Fisher**'s preference for experimental designs, delaying acceptance for decades. The revival came through two parallel traditions that would eventually reunite.

The **psychometric-econometric convergence** occurred in the 1970s. **Karl Jöreskog** developed maximum likelihood estimation for confirmatory factor analysis at ETS in 1967, then created LISREL (Linear Structural Relations) with **Dag Sörbom** at Uppsala University. The Jöreskog-Keesling-Wiley model (independently proposed 1972-73) unified econometric simultaneous equations, psychometric factor analysis, and sociological path models. The 1970 Social Science Research Council conference, published as Goldberger & Duncan 1973, marked the milestone bringing economists, psychologists, sociologists, and statisticians together. (PubMed Central)

Simultaneously, **Nelder and Wedderburn's 1972 GLM paper** showed that maximum likelihood estimates for a broad class of regression models—normal regression, ANOVA, logistic regression, log-linear models—could be obtained through iteratively weighted least squares. Their unification revealed three key components: a random component (exponential family distribution), systematic component (linear predictor $\eta = X\beta$), and link function $g(\mu) = \eta$.

The **mixed models tradition** developed separately through **Henderson's BLUP** for animal breeding (1950s) and the landmark **Laird & Ware 1982** paper establishing modern longitudinal mixed models with their classic formulation: $y_i = X_i\beta + Z_i b_i + \varepsilon_i$. **Breslow & Clayton 1993** introduced penalized quasi-likelihood for practical GLMM inference, merging GLMs with random effects.

The critical synthesis came with **Rabe-Hesketh, Skrondal, and Pickles' GLLAMM framework** in their 2004 *Psychometrika* paper and accompanying book *Generalized Latent Variable Modeling*. Their key insight: random intercepts are latent factors with loadings fixed to 1, random slopes are latent factors with covariate-dependent loadings, and factor structures generalize simple random effects to allow flexible covariance patterns. Muthén articulated this explicitly in 2002: "The idea of latent variables captures a wide variety of statistical concepts, including random effects, missing data, sources of variation in hierarchical data, finite mixtures, latent classes, and clusters."

**DSEM emerged in 2018** through Asparouhov, Hamaker, and Muthén's *Structural Equation Modeling* paper, integrating time-series dynamics (VAR models), multilevel structure (person-specific random effects), and SEM measurement models ⬭Springer for intensive longitudinal data like experience sampling and ecological momentary assessment studies. ⬭arXiv ⬭ResearchGate

## How Bayesian hierarchical modeling dissolves traditional distinctions

The Bayesian perspective, articulated most influentially by **Andrew Gelman** and **Richard McElreath**, reveals that the GLM/SEM distinction is artificial. Both involve latent (unobserved) parameters varying across units, population distributions describing this variation, observation models linking parameters to data, and prior information constraining estimation.

Gelman's seminal contribution frames random effects as existing on a **partial pooling continuum**: complete pooling ignores group structure, no pooling ignores shared information, and partial pooling draws group parameters from a common distribution. ⬭PyMC The partially-pooled estimate becomes a precision-weighted average of the group-specific mean and overall mean—groups with more data shrink less toward the grand mean, groups with less data borrow more strength from others. ⬭columbia

McElreath's *Statistical Rethinking* (winner of the 2024 De Groot Prize) emphasizes the **generative modeling philosophy**: specify beliefs about how data are generated rather than selecting techniques. His term "varying effects" over "random effects" captures this: "We define effects (or coefficients) in a multilevel model as constant if they are identical for all groups in a population and varying if they are allowed to differ from group to group."

The mathematical equivalence is concrete. A **confirmatory factor analysis** is a hierarchical model where indicators depend on latent factors. A **random intercepts model** is a hierarchical model where observations depend on group-specific intercepts. A **latent growth model** bridges both: simultaneously a repeated-measures multilevel model and an SEM. In Bayesian computation, both random effects and latent variables are treated identically—as parameters sampled from conditional posteriors.

**Priors function as structural constraints** analogous to SEM identification. Traditional SEM fixes a factor loading or latent variance to set scale. Bayesian priors provide "soft" constraints: informative priors encode structural assumptions, shrinkage priors (horseshoe, LASSO-type) regularize while allowing exploration, and small-variance priors allow cross-loadings while constraining toward zero. Muthén & Asparouhov 2012 proposed "replacing parameter specifications of exact zeros with approximate zeros based on informative, small-variance priors"—a more theoretically honest approach than assuming exact zeros. (PubMed)

## Implementing Bayesian SEM and DSEM in the PyMC ecosystem

For practical implementation, the tool landscape offers distinct strengths. **PyMC v5+** provides full capability for building Bayesian SEM/DSEM from scratch through its probabilistic programming primitives, with an official CFA/SEM tutorial available. **blavaan** in R offers the gold standard for Bayesian SEM using lavaan syntax with Stan backend, (mc-stan) providing built-in Bayesian fit indices (BRMSEA, BCFI) and posterior predictive model checking. **brms** handles SEM-like multivariate models through its formula interface but lacks traditional fit indices. (Scottclaessens)

The core PyMC pattern for a measurement model specifies factor loadings, latent factors with LKJ covariance priors, and residual variances:

```python
with pm.Model() as cfa_model:
    # Factor covariance (LKJ prior for correlations)
    Psi_chol, Psi_corr, Psi_stds = pm.LKJCholeskyCov(
        "Psi", n=2, eta=2.0, sd_dist=pm.Exponential.dist(1.0))

    # Latent factors (non-centered parameterization)
    eta_raw = pm.Normal("eta_raw", 0, 1, shape=(N, n_factors))
    eta = pm.Deterministic("eta", pm.math.dot(eta_raw, Psi_chol.T))

    # Residual variances
    theta = pm.HalfNormal("theta", sigma=1, shape=n_items)

    # Likelihood
    mu = pm.math.dot(eta, Lambda.T)
    y_obs = pm.Normal("y_obs", mu=mu, sigma=theta, observed=Y_data)
```

**Non-centered parameterization is critical** for hierarchical models. Instead of sampling $\theta \sim \text{Normal}(\mu\_\theta, \sigma\_\theta)$ directly (which creates funnel geometries), sample $\theta\_raw \sim \text{Normal}(0, 1)$ and transform: $\theta = \mu\_\theta + \sigma\_\theta \times \theta\_raw$. (Readthedocs) This dramatically improves MCMC efficiency.

For **within/between decomposition** in multilevel models, the Bayesian approach offers an elegant solution: estimate person means as parameters rather than computing them from observed data. This "latent decomposition" avoids the biases from observed-mean centering identified by Lüdtke and others.

**Convergence diagnostics** require R-hat < 1.01 (ideally < 1.005), effective sample size (ESS) > 400 per chain for both bulk and tail, and checking for divergent transitions. ArviZ provides comprehensive diagnostics: `az.summary()` for tabular output, `az.plot_trace()` and `az.plot_rank()` for visual inspection, and `az.plot_energy()` for HMC-specific diagnostics. Posterior predictive checks via `az.plot_ppc()` verify model adequacy.

For model comparison, **LOO-CV** (leave-one-out cross-validation via Pareto-smoothed importance sampling) is preferred over WAIC, with Pareto-k values < 0.7 (ideally < 0.5) indicating reliable estimates.

## Integrating DAGs with Bayesian estimation for causal inference

The mathematical foundation connecting DAGs to Bayesian models is the **factorization theorem**: for a DAG G, the joint probability factorizes as $P(X_1,...,X_n) = \prod P(X_i|Pa(X_i))$, where $Pa(X_i)$ represents parents of node $X_i$. `Lei Mao's Log Book` This compact representation maps directly to Bayesian model structure—each node becomes a random variable, each edge defines conditional dependency, and each conditional probability becomes a likelihood with priors.

**Pearl's do-calculus** provides rules for converting interventional distributions $P(Y|do(X))$ into estimable quantities. `arXiv` The workflow separates: (1) identification via do-calculus producing a non-parametric estimand, and (2) Bayesian estimation with full uncertainty quantification.

**PyMC v5.8+ includes a native** `do()` **operator** for interventional reasoning:

```python
from pymc import do

# Define generative causal model
with pm.Model() as causal_model:
    z = pm.Normal("z", 0, 1)
    x = pm.Normal("x", mu=gamma_zx * z, sigma=1)
    y = pm.Normal("y", mu=beta_xy * x + gamma_zy * z, sigma=1)

# Simulate intervention do(X=1)
model_do_x1 = do(causal_model, {"x": np.ones(N)})

# Counterfactual prediction
with model_do_x1:
    pred_y = pm.sample_posterior_predictive(idata, var_names=["y"])
```

**CausalPy** (from PyMC Labs) provides higher-level interfaces for quasi-experimental methods: synthetic control, difference-in-differences, regression discontinuity, and instrumental variables—all with full Bayesian uncertainty quantification.

For LLM-based DAG generation, recent research shows LLMs achieve **97% pairwise accuracy on the Tübingen benchmark** (vs 83% for statistical methods). (arXiv) The practical workflow: (1) LLM generates DAG from variable descriptions and domain context, (2) validate DAG structure (acyclicity, expert review), (3) use DoWhy for identification, (4) translate to PyMC model programmatically, (5) estimate with full posterior inference.

**Priors encode causal assumptions.** Regularizing priors (horseshoe, half-Cauchy) that shrink toward homogeneous effects work well for treatment effect estimation. (PubMed Central) A critical finding from Hahn et al. 2020: naive regularization can corrupt treatment effect estimation when treatment correlates with confounders—including propensity scores in the model specification resolves this "regularization-induced confounding."

## DSEM-specific considerations for intensive longitudinal data

Dynamic SEM handles data from experience sampling, ecological momentary assessment, and daily diary studies—many repeated measures (T > 30-100) per person with dense temporal spacing. (Taylor & Francis Online +2) The framework decomposes observed scores into **between-person variance** (stable trait-like differences), **between-time variance** (systematic time-specific variations), and **within-person variance** (moment-to-moment fluctuations). (arXiv)

The **Random Intercept Cross-Lagged Panel Model** (RI-CLPM), developed by Ellen Hamaker, addresses a fundamental flaw in traditional CLPM: if stability is trait-like and time-invariant, autoregressive parameters fail to account for this, leading to erroneous causal conclusions. (PubMed) RI-CLPM separates person-level random intercepts (between-person) from within-person deviations following VAR dynamics. (Taylor & Francis Online)

**Prior specification for temporal parameters requires care:**

| Parameter | Recommended Prior | Rationale |
|---|---|---|
| AR coefficients | Normal(0, 0.5) | Encourages stationarity without hard constraint |
| Cross-lagged effects | Normal(0, 0.3-0.5) | Rarely exceed |
| Random effect SDs | Half-Cauchy(0, 2.5) | Weakly informative, critical for small N |
| Correlations | LKJ(η=2) | Slight shrinkage toward zero |

Stan documentation advises **against hard stationarity constraints**: "If the data are not well fit by a stationary model it is best to know this." (Stan) Weakly informative priors that encourage but don't enforce stationarity provide better diagnostics.

For **irregular time intervals**, continuous-time parameterization (via the ctsem package) yields interval-invariant estimates. The key insight: discrete-time AR parameters depend on measurement interval, making

them incomparable across studies with different designs. (Journal of Statistical Software) Continuous-time drift matrices avoid this.

**Sample size requirements** from the methodological literature: N > 100 persons with T > 50 timepoints yields excellent convergence and unbiased parameter recovery; N = 50-100 with T = 30-50 works with slight variance bias; N < 50 requires informative priors for stable estimation. Bayesian DSEM's key advantage over frequentist alternatives: MCMC handles many random effects where maximum likelihood becomes intractable. (PubMed Central)

## Practical recommendations for building your framework

For someone building a causal inference framework combining LLM-based DAG specification with Bayesian PyMC estimation, the path forward involves four components working together:

**DAG specification layer**: Use LLMs for initial DAG generation with variable descriptions and domain context, then validate structure with domain experts. Hybrid approaches like CausalFusion that combine statistical tests with LLM orientation resolution perform best. Store DAGs in standard formats (GML, DOT) for toolchain compatibility.

**Identification layer**: Use DoWhy to apply backdoor/frontdoor criteria and derive estimands. Don't skip this step—it ensures your estimation targets a valid causal quantity. DoWhy's refutation methods (random common cause, placebo treatment) provide automated assumption checking. (Pywhy)

**Estimation layer**: CausalPy for standard quasi-experimental designs; custom PyMC models for complex causal structures with latent variables or temporal dynamics. Always use non-centered parameterizations for hierarchical components. Target accept_rate=0.95 or higher for complex models.

**Uncertainty communication**: Report full posterior distributions, not just point estimates—this is the key differentiator from frequentist approaches in practical decision-making. Credible intervals, posterior predictive checks, and sensitivity analysis to prior specification provide honest uncertainty quantification.

The theoretical unification—random effects as latent variables, both subsumed by Bayesian hierarchical models—has practical consequences: you're no longer choosing between techniques but specifying generative processes. The question shifts from "should I use mixed models or SEM?" to "what distributional assumptions and dependency structure match my understanding of how these data arose?" PyMC's probabilistic programming framework makes this perspective computationally tractable.

## Conclusion: one framework, many manifestations

The century-long progression from Wright's path analysis through LISREL, GLM, GLMM, GLLAMM, and DSEM represents a gradual recognition that these methods share common foundations in **latent variable modeling**. The Bayesian perspective makes this unity explicit and computationally operational.

For practitioners, this means freedom from artificial methodological silos. A measurement model with temporal dynamics and hierarchical structure is simply a Bayesian hierarchical model with specific features—specify it

directly in PyMC rather than searching for the "right" technique. Priors encode your assumptions transparently rather than hiding them in default settings or identification constraints.

The integration with causal inference via DAGs completes the picture: specify your causal assumptions graphically, identify estimands formally, and estimate with full uncertainty quantification. LLMs can accelerate DAG specification; Bayesian methods ensure honest uncertainty propagation to final causal effect estimates.

What remains challenging is prior specification for complex models (especially small-sample DSEM), computational scaling for very large datasets, and the gap between methodological publications and accessible implementation guidance. But the theoretical framework is settled: there's one underlying machinery, and the traditional categories are merely different windows into it.