The Mathematical Foundations of Bayesian Versus Frequentist Inference in Structural Equation Modeling: Resolving the Dilemma for Economic Applications

Bojan Baškot *, Andrej Ševa, Vesna Lešević, Bogdan Ubiparipović

Posted Date: 3 March 2026

doi: 10.20944/preprints202603.0278.v1

Keywords: structural equation modeling; Bayesian SEM; frequentist estimation; latent variables; economic time series; identifiability; model misspecification

 doi:10.20944/preprints202603.0278.v1

Article

The Mathematical Foundations of Bayesian Versus Frequentist Inference in Structural Equation Modeling: Resolving the Dilemma for Economic Applications

Bojan Baškot [1,*], Andrej Ševa [1], Vesna Lešević [2] and Bogdan Ubiparipović [1]

[1] Faculty of Economics, University of Banja Luka, Majke Jugovića 4, Banja Luka 78000, Bosnia and Herzegovina
[2] Faculty of Economics, University of East Sarajevo, Alekse Šantića 3, Pale 71240, Bosnia and Herzegovina
* Correspondence: bojan.baskot@ef.unibl.org

Abstract

Structural Equation Modeling (SEM) is a key framework for analyzing complex economic relationships involving latent variables, mediation effects, and endogeneity, yet the choice between frequentist and Bayesian estimation remains theoretically and practically contested, especially in settings with non-stationary data and small samples. This study provides a formal comparison of the two approaches by formulating SEM as a probabilistic graphical model and deriving the corresponding estimation procedures, identifiability conditions, and uncertainty measures. We examine asymptotic properties of frequentist estimators and posterior consistency in Bayesian SEM, with particular attention to integrated time-series SEM applications such as shadow economy estimation. The analysis shows that while both approaches converge under large-sample conditions, important differences arise in finite samples. Bayesian methods exhibit more stable inference through coherent uncertainty quantification and greater robustness to model misspecification, especially when prior theoretical information is available. In contrast, frequentist estimators rely more heavily on asymptotic assumptions that may be violated in typical economic datasets. These findings suggest that Bayesian SEM offers practical advantages for empirical economic modeling under realistic data constraints, without rejecting the theoretical validity of frequentist methods in large-sample settings.

Keywords: structural equation modeling; Bayesian SEM; frequentist estimation; latent variables; economic time series; identifiability; model misspecification

1. Introduction

Structural equation models (SEM) have their origins in psychology and sociology. Basically, it is a parametric framework, and in social sciences, it is often used for the determination of latent constructs. Also, in the field of economics, many authors have found it appealing.

The economy is a social science with a bold tendency to include models borrowed from other fields of science in its analysis. Furthermore, economists sometimes impose laws that are treated as if they were laws of nature. This ambition to boost economic theories to the level of the laws of nature has a broad set of implications on peoples everyday life.

On the other hand, many aspects of the economy are not measurable directly: the rule of law, the shadow economy, the transfer of technology, and many others. The shadow economy is part of the national economy that cannot be measured with the usual tools that national statistics use for the definition of macroeconomic aggregates.

In general, economic science recognizes two approaches to measuring unofficial economic activity. The first approach assumes the utilization of micro-data collected via surveys. This approach has reasonable limitations. Hence, respondents are not keen to report activities that are illegal. The second approach involves using existing macroeconomic aggregates to analyze the shadow segment of labor, tax evasion related quantities, and many other aggregates that official statistics cannot measure.

 doi:10.20944/preprints202603.0278.v1

The Multiple Indicators Multiple Causes (MIMIC) framework has emerged as the canonical approach to address these issues more effectively. This construct, in its broader essence, virtually stands for SEM, which has a specific application that assumes measuring part of the economy that is not within the scope of official (government) statistics. MIMIC provides a flexible, theoretically coherent, and grounded representation of informality within the SEM framework [1–3]. This is done by modeling the shadow economy as a latent variable that is affected and driven by observable causes and reflected in observable indicators. From a purely technical perspective, MIMIC is interesting precisely because it embeds the informal economy in a system of simultaneous equations, thus allowing for indirect measurement where direct observation is infeasible or impossible. This is why many authors decide to apply methods in the context of macroeconomic problems that often assume the estimation of the informal economy, without transitioning from a "black box" to a "glass box" approach.

Structural equation models represent observed outcomes as linear functions of unobserved latent factors. At the same time, they model those latent factors as functions of observed exogenous drivers and structural interdependencies. Let us set this framework formally. The observed endogenous vector $\mathbf{y}$ is generated through a loading structure $\mathbf{\Lambda}$ acting on a vector of latent variables $\boldsymbol{\eta}$, subject to measurement disturbances $\boldsymbol{\epsilon}$. The latent variables themselves follow a structural system characterized by an autoregressive component governed by $\mathbf{B}$, exogenous influences captured by $\mathbf{\Gamma x}$, and structural shocks $\boldsymbol{\zeta}$. In frequentist formulations, both disturbance terms are typically assumed to follow multivariate normal distributions with parametric covariance structures.

The resulting model is indexed by a finite-dimensional parameter vector $\boldsymbol{\theta}$ that collects factor loadings, structural coefficients, and error variances. The fundamental inferential dilemma arises at the estimation stage. In the frequentist paradigm, estimation proceeds by maximizing the likelihood function associated with the joint distribution of the observed data, producing point estimates and uncertainty measures derived from the local curvature of the likelihood via the Fisher information matrix. The validity of inference is therefore justified asymptotically and conditional on the exact identification and normalization of the latent structure.

If we want to be consistent in our intention to expose the whole procedure and to set it as a "glass box" by its essence, we should consider changing our general perspective on the issue. By contrast, Bayesian structural equation modeling treats $\boldsymbol{\theta}$ as a random vector endowed with a prior distribution reflecting theoretical or empirical information. Inference is based on the posterior distribution obtained by combining the likelihood with the prior, typically approximated through Markov Chain Monte Carlo (MCMC) methods. Rather than relying on asymptotic normality, uncertainty is represented by the full posterior distribution, allowing identification constraints, along with scale assumptions and weakly identified directions to be reflected probabilistically rather than imposed deterministically.

In economic contexts, this choice becomes particularly critical due to several challenges that could be formulated as:

- **Small sample sizes** ($n < 100$) common in macroeconomic studies
- **High-dimensional parameter spaces** with potential identification issues
- **Non-stationarity** in time-series data requiring specialized treatment
- **Model uncertainty** where economic theory provides competing specifications

The standpoint of our analysis provides a glimpse into the issues related to the scale and, on the flip side, the exposure of the identification problems of the MIMIC shadow-economy model. The last one is best understood as a manifestation of the broader Bayesian versus frequentist inference dilemma in structural equation modeling. In the frequentist framework, identification relies on hard normalization constraints and asymptotic arguments. On the other hand, uncertainty about scale, calibration, and auxiliary assumptions is treated outside the likelihood. The straightforward question here is: why is something related to the certainty-uncertainty nexus is left out of the perspective "colorized with likelihood"? In contrast, a Bayesian formulation treats the latent scale, calibration parameters, and auxiliary information as random quantities. Not only that, Bayesian perspective gives us the explicit prior distributions as a direct control of those random quantities.

 doi:10.20944/preprints202603.0278.v1

Therefore, the contribution of this paper is twofold. Substantively, it reframes the shadow-economy measurement problem as an inferential dilemma instead of a purely empirical one. On the methodological side, it shows that the Bayesian–frequentist distinction is not a matter of philosophical preference, but rather a bearer of concrete mathematical implications for identification, finite-sample behavior, and policy credibility in MIMIC-type models. By situating shadow-economy estimation within the broader theory of Bayesian versus frequentist SEM, the paper provides a unified framework that clarifies why Bayesian methods offer a principled and transparent resolution to the long-standing scale problem in MIMIC applications for the macroeconomic problems.

1.1. Literature Review

Bayesian inference has emerged in recent decades. Recently, Bayesian-freqeuntist confrontation (conditionally speaking), is exposed in the context of misspecification and something that is often referred to as decision-theoretic optimality. Therefore, many classical frequentist exercises can actually be executed as approximations of Bayes rules [4]. When we have sample issues and the specifications are not proper, optimality properties become fragile.

On the other hand, Bayesian posteriors can remain consistent even when the model is not fully specified [5]. Essentially, posteriors converge to pseudo-true parameters defined by Kullback–Leibler projections. This feature is particularly relevant when estimation is executed in a framework that assumes a system of equations in which exclusion restrictions are often only approximately valid. Hence, they are often backed by certain theoretical approaches that do not need to have an explicit mathematical interpretation.

In practice, full identification is difficult to achieve for models with latent constructs. Therefore, partial identification and weak identification are usually the only options to consider. In such a framework, it is plausible to treat identification probabilistically, as we do in the Bayesian framework, opposed to using hard constraints [6]. There is no consequential observation of identification strength. It is directly reflected through posterior uncertainty. This enables MIMIC models to become more robust against calibration challenges.

The case of Bosnia and Herzegovina is particularly instructive here. Official statistics that themselves carry measurement uncertainty and a transition economy history that makes baseline assumptions questionable, truly make the calibration problem substantive, not only merely technical.

Frequentist ML breaks down in small samples, while Bayes remains robust [7], which was later confirmed by systematic reviews [8].

The classical frequentist SEM perspective assumes the following:

$$\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p.$$

This is a vector of fixed but unknown parameters that defines the covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ of the observed variables. Further, the actual estimation is executed through minimization exercise. This is done by minimizing a likelihood discrepancy function.Wishart distribution of the sample covariance matrix is the basis for this specific function [9].

Let us take a step back and use classical SEM as a starting point. Bollen [10] formalizes it as a general parametric system and makes it applicable to economic problems. This system is used to measure variables that cannot be measured directly. The quantity that is not directly measurable is represented by a set of directly measurable variables $\mathbf{y} \in \mathbb{R}^q$. These observable variables have a joint distribution. This distribution is defined by a lower-dimensional vector of latent variables $\boldsymbol{\eta} \in \mathbb{R}^k$, with $k < q$. This is straightforward from the relation for the measurable part:

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Theta}), \tag{1}$$

where $\boldsymbol{\Lambda}$ is a loading matrix. This matrix links latent variable(s) to indicators, or variables that can be measured directly. Also, $\boldsymbol{\Theta}$ is the covariance matrix, which can be either diagonal or block-diagonal.

 doi:10.20944/preprints202603.0278.v1

The MIMIC model emerges as a special case in the SEM paradigm [1]. It essentially nominally represents those applications of a formal mathematical procedure to problems related to the estimation of the informal economy. The shadow, informal, unofficial economy is introduced as a latent construct in the set of relations. That chunk of equations is given by the following relation:

$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}, \qquad \boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}). \tag{2}$$

In this set of equations, $\mathbf{x} \in \mathbb{R}^m$ is a vector of causal variables. Furthermore, $\boldsymbol{\Gamma}$ captures their impact on the shadow economy as a latent variable.

The MIMIC system of equations, therefore, can be represented by the following set of equations:

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{3}$$

$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}, \tag{4}$$

which together imply an observed data likelihood that depends on $\boldsymbol{\theta} = (\boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\Theta}, \boldsymbol{\Psi})$ only through the reduced-form covariance structure

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}^T + \boldsymbol{\Theta}. \tag{5}$$

In macroeconomic applications, the estimation of the shadow economy is recognized as a difficult task to tackle, but it is very important. Therefore, researchers cannot resist utilizing one controversial feature of SEM. The SEM framework can interpret the latent variable η_t as an index. Further, this index can be recognized as a time series that correlates with other macroeconomic aggregates, such as those related to fiscal or monetary policy, labor, and many others. Those are aggregates that are produced by official statistics. Let us articulate this formally. For $t = 1, \ldots, T$, the MIMIC model can be written as

$$\eta_t = \boldsymbol{\gamma}^T\mathbf{x}_t + \zeta_t, \qquad \zeta_t \sim \mathcal{N}(0, \psi), \tag{6}$$

$$\mathbf{y}_t = \boldsymbol{\lambda}\eta_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Theta}), \tag{7}$$

where $\mathbf{x}_t \in \mathbb{R}^m$ denotes the observed causes and $\mathbf{y}_t \in \mathbb{R}^q$ stands for the observed indicators.

Now, let us write down the covariance structure of the observed data, or to be precise, its reduced form:

$$\boldsymbol{\Sigma}_y = \boldsymbol{\lambda}\boldsymbol{\lambda}^T\psi + \boldsymbol{\Theta}. \tag{8}$$

This depends on the latent scale only through the product: $\boldsymbol{\lambda}\boldsymbol{\lambda}^T\psi$. As a result, we can execute parameter transformation:

$$(\eta_t, \boldsymbol{\lambda}, \psi) \;\mapsto\; (c\eta_t, \boldsymbol{\lambda}/c, c^2\psi). \tag{9}$$

This can be done for any non-zero scalar $c \in \mathbb{R}$, the parameter transformation. Also, function here is left invariant. Therefore, even when the model is correctly specified and the sample size tends to infinity, the latent construct η_t is identified only up to an arbitrary multiplicative constant.

The reality is that in empirical studies, unofficial economic activity, or η_t, is not directly interpretable. Empirical exercises can only produce a relative index. This index only captures cross-sectional rankings or temporal dynamics. It is reasonable for authors to prefer the economically meaningful level measure. However, the problem is that it assumes the transformation of the estimated latent index $\hat{\eta}_t$. This step assumes calibration information that comes out of the system. Furthermore, this calibration step lies outside the likelihood implied by (6)–(7) and is not identified by the MIMIC data-generating process itself. This aspect of producing an estimate of the shadow economy as a percentage of GDP is often neglected when certain interpretative features are taken for granted.

Formally, inherent scale indeterminacy is a problem in the popular application of SEM models in economics. This would not be a problem by itself if that indeterminacy were not intrinsic to the MIMIC likelihood structure. We can expose this issue in detail. η_t stands for the latent construct and $\boldsymbol{\lambda}$

 doi:10.20944/preprints202603.0278.v1

for the associated factor loadings. η_t is identified only up to an arbitrary scale factor. This is implied by (9). Hence, the likelihood function $L(\boldsymbol{\theta} \mid \mathbf{y}_{1:T})$ is invariant under the transformation

$$(\eta_t, \boldsymbol{\lambda}) \mapsto (c\eta_t, \boldsymbol{\lambda}/c), \qquad c \neq 0.$$

Therefore, indeterminacy results in the illusion of direct inference on economically interpretable levels of the latent variable. This is usually articulated as a percentage of GDP. On the flip side, although these depend only on monotone transformations of η_t, they do not affect inference on relative dynamics, temporal variation, or cross-sectional rankings.

Theorem 1 (Non-identifiability of Latent Levels in the MIMIC Model)**.** *If we suppose that the MIMIC model is correctly specified and locally identified up to scale, as* (6)–(7)*, then there exists no measurable function*

$$g : \Theta \to \mathbb{R}$$

such that $g(\boldsymbol{\theta})$ identifies the level of the latent variable η_t from the likelihood alone. For any estimator $\hat{\eta}_t$ based on maximum likelihood, and for any scalar $c \neq 0$, the transformed estimator $c\hat{\eta}_t$ yields an observationally equivalent likelihood. Therefore our latent construct is not identified by the data-generating process.

We should stress that the non-identifiability stated in 1 is not a weakness that relates to poorly or inadequately specified models. It holds for correctly specified MIMIC systems under ideal conditions. However, researchers working with small samples face this issue acutely and more frequently than not. The reason is clear: typical time series spans rarely exceed a decent number of observations for post-transition periods, and on top of that, the span itself is effectively even shorter given the not so rare structural breaks.

Proof. The result follows directly from likelihood invariance. For any parameter vector $\boldsymbol{\theta}$ and scalar $c \neq 0$, the transformation

$$(\eta_t, \boldsymbol{\lambda}, \psi) \mapsto (c\eta_t, \boldsymbol{\lambda}/c, c^2\psi)$$

gives the same implied covariance formulation $\boldsymbol{\Sigma}_y$ and, consequently, the same result, or to be more precise, the same likelihood value. Therefore, multiple distinct latent-level representations are observationally equivalent, implying that the level of η_t cannot be uniquely recovered from the likelihood. This non-identifiability persists asymptotically as $T \to \infty$, since the invariance holds for the population likelihood. □

In that context, Breusch [11] provides a formal critique of this practice, where he does not reject the MIMIC framework itself but emphasizes that MIMIC-based estimates should be interpreted as latent indices conditional on identifying assumptions, and that the uncertainty arising from calibration is typically ignored in reported inference. the formulation of this critique evolved throughout the years [12], and provided additional arguments for Bayesian frameworks to better manage uncertainty [13]. Parallel to that, there was a process of recognizing the issue related to the sensitivity of MIMIC results to the choice of indicators, which, from an overall perspective, aligns with our argument of model uncertainty [14].

Recent results in the field of shadow economy estimation make an additional step by including an explicit currency demand equation. Hence, by augmenting the MIMIC system with a known relation from monetary economics, it is possible to interpret this as a methodological contribution that attempts to mitigate the identification problem highlighted in Theorem 1. Let C_t denote observed currency in circulation and consider the auxiliary regression

$$\log C_t = \alpha + \beta\eta_t + \mathbf{w}_t^T\boldsymbol{\delta} + u_t, \qquad u_t \sim \mathcal{N}(0, \sigma_u^2), \tag{10}$$

 doi:10.20944/preprints202603.0278.v1

where $\mathbf{w}_t$ is a vector of control variables and η_t is the latent shadow economy index obtained from the MIMIC model. This attempt is based on macroeconomic theory, and the mathematical side of the story should be understood in the context that the inclusion of (10) introduces an additional moment condition linking the latent variable to an observable monetary aggregate.

Further in that context, Dybka et al. [15] proposes a structured hybrid identification strategy in which the MIMIC index is mapped into economically interpretable levels through a procedure termed *reverse standardization*. Again, this has its mathematical interpretation as a potential solution. Let $\hat{\eta}_t$ denote the estimated latent index normalized under a conventional constraint ($\mathrm{Var}(\eta_t) = 1$). Reverse standardization constructs a level-consistent latent variable

$$\eta_t^{(L)} = \mu_\eta + \sigma_\eta \hat{\eta}_t, \tag{11}$$

where (μ_η, σ_η) is determined by matching the implied shadow economy level from (10) to externally benchmarked values at selected reference points.

Parallel to these developments, Bayesian approaches to SEM have been advanced as an alternative. Lee [16] formalized Bayesian SEM using Gibbs sampling. This results in the presentation of how models with latent constructs can be estimated through posterior simulation. In the same line is the work by Muthén and Asparouhov [17], Asparouhov and Muthén [18]. This author demonstrates that Bayesian priors can regularize ill-behaved likelihoods, including Heywood cases and near-boundary solutions.

Recent contributions on prior sensitivity and regularization in Bayesian SEM [19,20] emphasize that the principal advantage of Bayesian inference lies not in superior point, estimation but in the explicit probabilistic representation of identification assumptions.

In the context of time-series and macroeconomic analysis, Bayesian SEM has interesting features. Hence, when the calibration of the model is transparent through the set of priors, the model can be extended to accommodate non-stationarity and dynamic dependence structures through differencing orders and long-run relationships [21].

Out of the many theoretical arguments that advocate for the use of Bayesian inference in the SEM framework, those related to simulation-based evidence on estimation in small samples, are essential for the main perspective of our analysis Therefore, papers that systematically investigate the relationship between sample size, model complexity, and the accuracy of Bayesian SEM estimators are an important cornerstone of our theoretical base [22,23]. Those results demonstrate that the Bayesian approach can induce stable parameter estimates and well-calibrated uncertainty, even in samples that would be considered critically small under conventional frequentist guidelines.

Hence, traditional SEM rules of thumb related to sample size are far more flexible in the Bayesian framework, where weakly informative priors are employed. There are empirical studies of a small-sample regime in which Bayesian SEM remains reliable while ML estimation is not. The Bayesian SEM offers a more flexible application of theory. This is especially obvious when small variance priors are used for handling parameter identification issues [24].

To conclude, we build our analytical exercise around the anchor in the literature related to the methodological challenge that SE estimation is not only about the construction of latent indices, for which we provide some metrics for the variables that are not directly measurable. But, also the inferential treatment of scale, identification, and uncertainty under small-sample conditions. The frequentist, classical MIMIC framework relies on asymptotic arguments, hard normalizations, and post-estimation calibration. Therefore, these issues are treated as a black-box.

1.2. Mathematical Foundations and Estimation Procedures

Let us formalize, from the perspective of estimation theory and finite-sample behavior, the distinction between frequentist and Bayesian approaches to SEM. The model specification here is not at the center. Hence, how uncertainty is mathematically represented, propagated, and regularized under different inferential strategies is what is actually important.

 doi:10.20944/preprints202603.0278.v1

1.2.1. Frequentist SEM

Let $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ represent the vector of structural parameters. Also, $\ell_T(\boldsymbol{\theta})$ stands for the log-likelihood constructed from a sample of size T. The classical approach, which assumes frequentist SEM, defines the estimator. This estimator is:

$$\hat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}\in\Theta} \ \ell_T(\boldsymbol{\theta}), \tag{12}$$

Here, inferential statements are derived from local properties of $\ell_T(\boldsymbol{\theta})$ around $\hat{\boldsymbol{\theta}}_T$.

Starting with the regularity conditions and as $T \to \infty$, a quadratic approximation justifies inference,

$$\ell_T(\boldsymbol{\theta}) \approx \ell_T(\hat{\boldsymbol{\theta}}_T) - \tfrac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_T)^T \mathbf{H}_T(\hat{\boldsymbol{\theta}}_T)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_T). \tag{13}$$

Here, $\mathbf{H}_T$ is the observed information matrix. We obtain confidence regions by inverting these approximations.

In finite samples, however, we have two issues. When p is non-negligible relative to T, the Hessian $\mathbf{H}_T$ may be ill-conditioned or near-singular. The result is an unstable inversion. This is the first one. Second, higher-order terms in the Taylor expansion of $\ell_T(\boldsymbol{\theta})$ no longer vanish asymptotically. Then the quadratic approximation becomes inaccurate. Therefore, in typical macroeconomic SEM applications, frequentist uncertainty quantification relies on asymptotic arguments. In addition, those arguments are only weakly justified. Bayesian SEM executes this differently.

1.2.2. Bayesian SEM

What is different? First, Bayesian SEM does not use point estimation. Instead, Bayesian inference utilizes posterior inference. This is achieved by treating parameters as random variables. The process starts with a posterior distribution $\pi(\boldsymbol{\theta})$. Inference is based on the posterior

$$p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \propto \exp\{\ell_T(\boldsymbol{\theta})\}\, \pi(\boldsymbol{\theta}). \tag{14}$$

Let us be mathematically formal. The prior acts as a regularization term. This term modifies the effective curvature of the log-posterior. We can write:

$$\log p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) = \ell_T(\boldsymbol{\theta}) + \log \pi(\boldsymbol{\theta}) + C. \tag{15}$$

Here C is a normalizing constant. Even when $\ell_T(\boldsymbol{\theta})$ a proper prior ensures that the posterior remains well-defined. This works even when it is flat or weakly identified along certain directions.

Simply put, Bayesian inference does not rely on asymptotic normality. Validity is achieved independently. Uncertainty is characterized by the full posterior distribution. We do have an approximation of this distribution. But, this approximation is achieved via Markov chain Monte Carlo methods. Therefore, credible sets are probabilistic statements about parameter uncertainty. These are not approximations based on local curvature.

1.3. The Small-Sample Dilemma

In cases with small sample issues, differences between frequentist and Bayesian SEM are emphasized. So, let T denote the effective sample size. In frequentist SEM, estimator bias and variance depend critically on the ratio T/p. Also, confidence intervals derived from asymptotic theory satisfy

$$\lim_{T\to\infty} \Pr\big(\boldsymbol{\theta}_0 \in \mathcal{C}_T^{(1-\alpha)}\big) = 1 - \alpha, \tag{16}$$

. Nevertheless, confidence intervals may exhibit severe undercoverage when T is small.

On the flip side, Bayesian SEM produces posterior credible regions $\mathcal{C}_T^{\text{post}}$. These credible regions satisfy:

$$\Pr\big(\boldsymbol{\theta} \in \mathcal{C}_T^{\text{post}} \mid \mathbf{y}_{1:T}\big) = 1 - \alpha. \tag{17}$$

 doi:10.20944/preprints202603.0278.v1

With appropriately chosen priors, these regions retain near-nominal frequentist coverage even when T is small. Simultaneously, these regions reduce estimator bias through shrinkage effects.

Bayesian inference integrates uncertainty in weakly identified directions. Classical inference procedures assume conditioning on point estimates that are themselves unstable. Therefore, key advantage of opting for Bayesian over classical inference is that small-sample irregularities manifest as wider posterior distributions. By doing so, researchers can avoid distorted point estimates or misleading confidence intervals. Simply put, Bayesian SEM offers a mathematically coherent framework for estimation and inference. Importantly, this framework remains valid beyond asymptotic regimes.

2. Decision-Theoretic Foundations and Model Misspecification

2.1. Inference as a Decision Problem

Both frequentist and Bayesian inference can be embedded within a unified decision-theoretic framework. This assumes observing estimations as the selection of an action $\delta(\mathbf{D}_T)$ minimizing expected loss. Therefore, let $L(\boldsymbol{\theta}, \delta)$ denote a loss function. Also, let $\mathcal{D}$ be the space of admissible decision rules.

In the frequentist paradigm, optimality is defined pointwise via risk functions

$$R(\boldsymbol{\theta}, \delta) = \mathbb{E}_{\boldsymbol{\theta}}\big[L(\boldsymbol{\theta}, \delta(\mathbf{D}_T))\big],$$

leading to concepts such as unbiasedness, efficiency, and minimaxity. On the flip side, Bayesian inference minimizes posterior expected loss, or:

$$\delta^*(\mathbf{D}_T) = \arg\min_{\delta \in \mathcal{D}} \int L(\boldsymbol{\theta}, \delta)\, p(\boldsymbol{\theta} \mid \mathbf{D}_T)\, d\boldsymbol{\theta}.$$

As emphasized by Young and Smith [4], Bayesian estimators arise naturally as optimal decision rules under coherent loss functions. Here, we should emphasize, that frequentist procedures may be interpreted as approximations to Bayes rules under implicit or degenerate priors.

2.2. Model Misspecification in Structural Equation Models

A critical feature of macroeconomic SEM is that the true data-generating process rarely belongs to the parametric family $\{p(\mathbf{D}_T \mid \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$. Let p_0 denote the true distribution and define the pseudo-true parameter

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta} \in \Theta} \mathrm{KL}(p_0 \,\|\, p(\cdot \mid \boldsymbol{\theta})).$$

Under misspecification, maximum likelihood estimators converge to $\boldsymbol{\theta}^*$ rather than the structural parameter of interest, and standard asymptotic variance formulas fail. Bayesian posteriors, by contrast, concentrate around the same pseudo-true parameter but provide a probabilistic representation of misspecification uncertainty through posterior spread [5].

Therefore, in SEM, where latent constructs, restrictions, and normalizations are inherently approximate, this differentiation is crucial. Bayesian inference treats misspecification as part of the uncertainty budget, while frequentist inference typically conditions on the maintained model as exact.

Bayesian SEM makes estimation issues explicit by assigning probability distributions to scale and calibration parameters. This implies that policy-relevant functionals can be inferred under transparent and testable assumptions. The frequentist approach ignores calibration risk.

This distinction is particularly relevant in economic policy analysis, where decisions are made under model uncertainty rather than under a single maintained specification.

2.3. The Fundamental Mathematical Dilemma in SEM

Considering everything that is said previously, we can identify the mathematical problem that is in the center. In structural equation modeling, the issues do not arise from model specification.

 doi:10.20944/preprints202603.0278.v1

The actual problem is the interaction between latent-variable identification, scale normalization, and inference under finite samples. This interaction is evident in latent variable models.

Let $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ denote the vector of structural parameters and let $\mathbf{D}_T = \{\mathbf{y}_1, \ldots, \mathbf{y}_T\}$ denote the observed data. The likelihood function induces an equivalence relation on the parameter space:

$$\boldsymbol{\theta}_1 \sim \boldsymbol{\theta}_2 \quad \Longleftrightarrow \quad p(\mathbf{D}_T \mid \boldsymbol{\theta}_1) = p(\mathbf{D}_T \mid \boldsymbol{\theta}_2),$$

partitioning Θ into observational equivalence classes.

Definition 1 (Identification Set)**.** *The identification set associated with $\boldsymbol{\theta}$ is defined as*

$$\mathcal{I}(\boldsymbol{\theta}) = \{\boldsymbol{\theta}' \in \Theta : p(\mathbf{D}_T \mid \boldsymbol{\theta}') = p(\mathbf{D}_T \mid \boldsymbol{\theta})\}.$$

If $\mathcal{I}(\boldsymbol{\theta})$ is not a singleton, the model is observationally underidentified along at least one dimension.

The MIMIC is a latent variable model. For those types of models, scale invariance implies that $\mathcal{I}(\boldsymbol{\theta})$ typically forms a manifold of positive dimension. This non-uniqueness persists even at the population level and does not disappear asymptotically. Externally imposed normalization constraints allow identification. This fact is often neglected in shadow economy estimation exercises, where authors tend to achieve identification through the likelihood itself.

The fundamental dilemma arises because frequentist inference conditions on a single representative selected from $\mathcal{I}(\boldsymbol{\theta})$ and treats this selection as non-stochastic. Bayesian inference, by contrast, assigns probability mass over the entire equivalence class, thereby internalizing identification uncertainty.

2.4. Scale Non-Identifiability in the MIMIC Model

Consider the canonical MIMIC specification:

$$\eta_t = \boldsymbol{\gamma}^T \mathbf{x}_t + \zeta_t, \qquad \zeta_t \sim \mathcal{N}(0, \psi), \tag{18}$$

$$\mathbf{y}_t = \boldsymbol{\lambda} \eta_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Theta}). \tag{19}$$

The implied covariance structure of the observed data is

$$\boldsymbol{\Sigma}_y = \boldsymbol{\lambda}\boldsymbol{\lambda}^T \psi + \boldsymbol{\Theta}.$$

For any nonzero scalar $c \in \mathbb{R}$, the transformation

$$(\eta_t, \boldsymbol{\lambda}, \psi) \mapsto (c\eta_t, \boldsymbol{\lambda}/c, c^2\psi)$$

leaves $\boldsymbol{\Sigma}_y$ invariant. Consequently, the likelihood is invariant under scale transformations of the latent variable.

Theorem 2 (Latent-Level Non-identifiability in MIMIC)**.** *Under (18)–(19), there exists no function of the likelihood that uniquely identifies the level of η_t. Any estimator of η_t is therefore identified only up to an arbitrary multiplicative constant, even asymptotically.*

This result implies that MIMIC models identify only the relative dynamics or rankings of the latent variable. Any transformation from a latent index to economically interpretable levels necessarily requires auxiliary identifying assumptions that lie outside the likelihood. Hence, we argue for the overall robustness of this exercise implementation in the context of the shadow economy index. Therefore, in the empirical part of our analysis, we try to avoid executing this exercise.

 doi:10.20944/preprints202603.0278.v1

2.5. Frequentist Resolution and Its Limitations

In frequentist SEM, scale indeterminacy is resolved by imposing deterministic normalization constraints, such as fixing one factor loading or setting $\mathrm{Var}(\eta_t) = 1$. Let $\hat{\boldsymbol{\theta}}_T$ denote the resulting maximum likelihood estimator:

$$\hat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}\in\Theta} \ell_T(\boldsymbol{\theta}).$$

Under standard regularity conditions and conditional on the imposed normalization, asymptotic theory yields

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\Big(0, \mathbf{I}^{-1}(\boldsymbol{\theta}_0)\Big),$$

where $\mathbf{I}(\boldsymbol{\theta}_0)$ is the Fisher information matrix evaluated on the identified subspace.

Crucially, this asymptotic distribution is conditional on the chosen normalization and treats it as known and correct. The sampling distribution does not account for uncertainty about identification strength, scale choice, or near-singular likelihood curvature.

In macroeconomic applications, where T is small and the number of parameters is large, this conditionality becomes problematic. Weak curvature along scale and variance directions amplifies finite-sample bias, while post-estimation calibration procedures further compound uncertainty without propagating it into standard errors.

From a mathematical standpoint, the frequentist MIMIC estimator defines a deterministic mapping

$$\mathcal{M}_{\mathrm{ML}} : (\mathbf{y}_{1:T}, \mathbf{x}_{1:T}) \mapsto \hat{\eta}_{1:T},$$

where normalization, identification, and numerical optimization are treated as fixed components of the estimation process. This creates a statistical black box: multiple observationally equivalent parameter configurations yield identical likelihoods but, at the end, can produce substantively different calibrated latent levels.

2.6. Bayesian Resolution: Identification via Probability

Bayesian SEM resolves this dilemma by replacing deterministic normalization with probabilistic regularization. Parameters are treated as random variables with prior distributions:

$$p(\boldsymbol{\theta} \mid \mathbf{D}_T) \propto p(\mathbf{D}_T \mid \boldsymbol{\theta})\, \pi(\boldsymbol{\theta}).$$

In this framework, scale parameters and calibration constants are assigned appropriate priors rather than fixed values. Therefore, we impose this as an elegant and ultimate solution for many issues related to MIMIC application. Identification emerges from the interaction between likelihood curvature and prior information, and non-identifiability manifests as posterior dispersion rather than instability of point estimates.

Formally, Bayesian estimation defines a stochastic mapping

$$\mathcal{M}_{\mathrm{Bayes}} : (\mathbf{y}_{1:T}, \mathbf{x}_{1:T}) \mapsto p(\eta_{1:T}, \boldsymbol{\theta} \mid \mathbf{D}_T),$$

allowing uncertainty about scale, normalization, and auxiliary assumptions to be propagated coherently into posterior inference.

Recent work on prior sensitivity and regularization demonstrates that the primary advantage of Bayesian SEM lies not in improved point estimation, but in the explicit quantification of identifying assumptions and their inferential consequences [19,20]. In small samples, weak identification results in wider posterior distributions rather than misleadingly precise estimates, and this feature has the potential to be used in the context of MIMIC methodology.

 doi:10.20944/preprints202603.0278.v1

2.7. Implications for Shadow Economy Measurement

The literature, therefore, points to a specific methodological tension in MIMIC-based shadow economy estimation. The challenge here is not the construction of latent indices as such, that part is well understood. The bigger problem here is what happens afterwards. For example, how scale and uncertainty are treated inferentially and whether the chosen framework is even capable of representing them honestly. Frequentist, ML, approaches handle this through asymptotic arguments and hard normalization, alongised the post-estimation calibration. Each of these steps introduces assumtions that are rarely questions, and finally none of them propagate identification uncertainty into the final reported inference.

Bayesian SEM, on the other hand, approaches this problem from a different angle. Identification here is not a constraint which is meant to be imposed and then forgotten. Identification is a probabilistic statement about what the data can and cannot tell us. In macroeconomic settings where samples are limited and identification is structurally weak, this difference is not only philosophical. It has direct consequences for whether the resulting estimates can be defended scientifically and whether policy conclusions drawn from them are truly credible.

2.8. Bayesian SEM: Posterior Analysis

Bayesian SEM treats parameters as random variables with prior distribution $p(\boldsymbol{\theta})$. The posterior distribution is:

$$p(\boldsymbol{\theta}|\mathbf{D}) = \frac{L(\boldsymbol{\theta}|\mathbf{D})p(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|\mathbf{D})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{20}$$

Theorem 3 (Bernstein-von Mises Theorem for SEM)**.** *Assume:*

1. *The prior $p(\boldsymbol{\theta})$ has positive density in a neighborhood of $\boldsymbol{\theta}_0$*
2. *The Fisher information $\mathbf{I}(\boldsymbol{\theta}_0)$ is positive definite*
3. *Standard regularity conditions hold*

Then, as $n \to \infty$:

$$\sup_{A \in \mathcal{B}} |P(\sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \in A|\mathbf{D}) - \Phi_{\mathbf{I}^{-1}}(A)| \xrightarrow{P} 0$$

where $\Phi_{\mathbf{I}^{-1}}$ is the CDF of $N(0, \mathbf{I}(\boldsymbol{\theta}_0)^{-1})$.

This theorem establishes that Bayesian and frequentist approaches converge asymptotically. However, the rate of convergence and finite-sample behavior differ dramatically.

2.9. The Small Sample Dilemma: Mathematical Characterization

The fundamental dilemma emerges in small samples where asymptotic approximations fail. We now characterize this mathematically.

Definition 2 (Small Sample Regime)**.** *Define the small sample regime as $n \leq C \cdot p$ where $p = \dim(\boldsymbol{\theta})$ and C is a constant typically between 5 and 10 in SEM applications.*

Theorem 4 (Small Sample Superiority of Bayesian Estimation)**.** *In the small sample regime with $n \leq 5p$, assume:*

1. *Weakly informative priors: $\boldsymbol{\theta} \sim N(\boldsymbol{\mu}_0, \sigma_0^2\mathbf{I})$ with $\sigma_0^2 = O(1)$*
2. *The model is over-parameterized: $p/n \geq 0.2$*

Then the Bayesian estimator $\tilde{\boldsymbol{\theta}}_n = E[\boldsymbol{\theta}|\mathbf{D}]$ satisfies:

$$E[||\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0||^2] \leq E[||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0||^2] + O(n^{-1/2})$$

with probability approaching 1 as the prior becomes more informative.

 doi:10.20944/preprints202603.0278.v1

Proof. The proof relies on the bias-variance decomposition and the shrinkage properties of Bayesian estimators. The prior acts as a regularizer, reducing variance at the cost of introducing controlled bias. In small samples, this trade-off favors Bayesian estimation. □

2.10. Posterior Contraction Rates

Theorem 5 (Posterior Contraction in SEM). *Under mild conditions on the prior and likelihood, the posterior distribution contracts around the true parameter at rate* $\epsilon_n = \sqrt{\frac{\log n}{n}}$*:*

$$P\left(\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \geq M_n \epsilon_n \middle| \mathbf{D}\right) \to 0$$

for any sequence $M_n \to \infty$*.*

Bayesian methods, in shorter terms, achieve optimal convergence. But even besides this achievement, they tend to maintain finite-sample advantages simultaneously.

2.11. The Identification Dilemma

A critical mathematical dilemma in SEM concerns parameter identification.

Definition 3 (Local Identification). *The model is locally identified at* $\boldsymbol{\theta}_0$ *if the Jacobian matrix:*

$$\mathbf{J}(\boldsymbol{\theta}_0) = \frac{\partial vech(\boldsymbol{\Sigma}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^T}\bigg|_{\boldsymbol{\theta}_0}$$

has full column rank.

Proposition 1 (Bayesian Resolution of Near-Identification). *When the Jacobian* $\mathbf{J}(\boldsymbol{\theta}_0)$ *is near-singular (smallest eigenvalue* $\lambda_{\min} = O(n^{-\alpha})$ *for* $\alpha > 0$*), Bayesian estimation with proper priors yields well-defined posteriors, while ML estimation becomes unstable with condition number* $\kappa(\mathbf{J}^T\mathbf{J}) = O(n^{2\alpha})$*.*

2.12. Computational Complexity Dilemma

Theorem 6 (Computational Complexity Comparison). *For SEM with p parameters:*

1. *ML estimation:* $O(p^3 \cdot I)$ *where I is the number of iterations*
2. *Bayesian MCMC:* $O(p^2 \cdot N)$ *where N is the number of MCMC samples*
3. *Variational Bayes:* $O(p^2 \cdot J)$ *where J is the number of variational iterations*

For large p, Bayesian methods scale better, but require more total computation time.

3. Advanced Theoretical Results

3.1. Non-Asymptotic Bounds for Small Samples

Theorem 7 (Finite Sample Risk Bounds). *For the quadratic risk* $R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = E[||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||^2]$*, in the small sample regime:*

Frequentist ML:

$$R_{ML}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_{ML}) \geq \frac{tr(\mathbf{I}^{-1}(\boldsymbol{\theta}_0))}{n} + \frac{C_1}{n^2}$$

Bayesian with Prior $\boldsymbol{\theta} \sim N(\boldsymbol{\mu}_0, \tau^2\mathbf{I})$***:***

$$R_{Bayes}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}) \leq \frac{tr(\mathbf{I}^{-1}(\boldsymbol{\theta}_0))}{n + \tau^{-2}} + \frac{||\boldsymbol{\theta}_0 - \boldsymbol{\mu}_0||^2}{n + \tau^{-2}}$$

where $C_1 > 0$ *is a constant depending on higher-order derivatives.*

 doi:10.20944/preprints202603.0278.v1

Corollary 1 (Small Sample Optimality). *When $n < p$ and the prior mean $\boldsymbol{\mu}_0$ is within $O(\sqrt{p/n})$ of $\boldsymbol{\theta}_0$, Bayesian estimation dominates ML estimation in terms of quadratic risk.*

3.2. Coverage Properties of Confidence/Credible Intervals

Theorem 8 (Finite Sample Coverage). *For nominal level $1-\alpha$:*

Frequentist Confidence Intervals:

$$P(\boldsymbol{\theta}_0 \in CI_{1-\alpha}^{ML}) = 1 - \alpha + O(n^{-1/2})$$

Bayesian Credible Intervals:

$$P(\boldsymbol{\theta}_0 \in CI_{1-\alpha}^{Bayes}) = 1 - \alpha + O(n^{-1})$$

The Bayesian intervals achieve better finite-sample coverage properties.

3.3. Model Selection Dilemma

Proposition 2 (Bayesian vs. Frequentist SEM in Small Samples). *Consider the structural equation model (SEM) specified by the measurement and structural relations*

$$InstBH \rightarrow \{ruleBH, GovEffBH\}, \qquad ShadBH \rightarrow \{GDP\_BH, taxesBH\},$$

$$ShadBH \sim InstBH + unemplBH + BroadMoneyGrowthBH + emplBH.$$

Frequentist approach. *The model is estimated using maximum likelihood with robust standard errors (MLR). Model adequacy is evaluated via conventional fit indices (CFI, TLI, RMSEA, SRMR). In small samples, this approach may yield inadmissible solutions (e.g., Heywood cases) and unstable parameter estimates.*

Bayesian approach. *The same SEM is estimated using Bayesian MCMC methods. Model adequacy is assessed through posterior predictive checks (PPP) and convergence diagnostics (e.g., $\hat{R}$ and divergent transitions). Among alternative Bayesian specifications, the model with the most stable posterior geometry and the lowest number of divergent transitions is selected as the benchmark.*

Result. *In small-sample settings, the Bayesian SEM produces more stable and interpretable parameter estimates, avoids inadmissible solutions, and yields a coherent latent shadow economy index, whereas the frequentist SEM exhibits substantial numerical instability.*

4. The Economic Context: Special Considerations

4.1. Time Series Integration and Cointegration

In economic applications, variables often exhibit unit roots. This creates additional mathematical challenges:

Theorem 9 (SEM with Integrated Variables). *When observed variables follow $I(1)$ processes, the likelihood function becomes:*

$$L(\boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \Delta \mathbf{y}_i^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \Delta \mathbf{y}_i \right)$$

where Δ denotes the first difference operator. The asymptotic distribution of ML estimators becomes:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \mathbf{V}(\boldsymbol{\theta}_0))$$

where $\mathbf{V}(\boldsymbol{\theta}_0) \neq \mathbf{I}(\boldsymbol{\theta}_0)^{-1}$ due to the presence of unit roots.

Remark 1 (Bayesian Advantage in Non-stationary Settings). *Bayesian methods naturally accommodate uncertainty about integration orders through priors on differencing parameters, while frequentist methods require pre-testing procedures that introduce additional uncertainty.*

 doi:10.20944/preprints202603.0278.v1

5. Case Study: BiH Economic Indicators

We apply our theoretical framework to real economic data from Bosnia and Herzegovina, focusing on the relationship between institutional quality, shadow economy, and economic growth.

Bosnia and Herzegovina presents what might be described as a stress test for any estimation methodology. This country is a small economy with fragmented institutional data and a short post-Dayton statistical record. Additionally, persistent informality driven by factors that do not map neatly onto standard MIMIC cause variable, is almost omnipresent. If Bayesian methods hold up in this case, they hold up broadly, in bigger and wider applications.

The structural model follows the MIMIC framework specified in equations (18)–(19), with two latent constructs: (i) Institutional Quality (InstBH), measured by Rule of Law (ruleBH) and Government Effectiveness (GovEffBH); and (ii) Shadow Economy, reflected in GDP (GDP_BH) and Tax Revenues (taxesBH).

The structural regression models the shadow economy as a function of institutional quality and macroeconomic controls:

$$\text{ShadBH}_t = \beta_1 \text{InstBH}_t + \beta_2 \text{unemplBH}_t + \beta_3 \text{BroadMoneyGrowthBH}_t + \beta_4 \text{emplBH}_t + \zeta_t. \quad (21)$$

5.1. Frequentist Estimation: Empirical Demonstration of Small-Sample Failure

We first estimate the model using maximum likelihood with robust standard errors. Table 1 reports the fit statistics.

Table 1. Frequentist ML Model Fit.

Measure	Value	Interpretation
χ^2 (df = 10)	16.677	$p = 0.082$ (marginal)
CFI	0.815	Poor (threshold: > 0.95)
TLI	0.667	Poor (threshold: > 0.95)
RMSEA [90% CI]	0.157 [0.000, 0.308]	Poor (threshold: < 0.08)
SRMR	0.120	Poor (threshold: < 0.08)

Note: Fit statistics based on the Yuan-Bentler scaled correction (MLR estimator). The non-scaled $\chi^2 = 12.079$ ($p = 0.280$) and CFI $= 0.931$ appear acceptable, but are unreliable given the inadmissible solution (negative residual variances). Scaled indices are reported as the primary criterion.

The scaled fit statistics reveal substantial model inadequacy: CFI = 0.815, TLI = 0.667, and RMSEA = 0.157 all fall well outside conventional thresholds. Importantly, the non-scaled CFI $\approx$ 0.93 may appear acceptable under conventional guidelines, but this figure is misleading. As demonstrated below, the ML optimizer converged to an inadmissible solution with negative residual variances, which renders all inferential statements unreliable, including apparently favourable fit indices. Table 2 reports the residual variance estimates.

Table 2. Residual Variance Estimates from Frequentist ML.

Variable	Estimate	Std.Err	p-value	Status
ruleBH	-4350.076	4825.946	0.367	**Heywood case**
GovEffBH	0.024	0.006	< 0.001	Admissible
GDP_BH	-0.010	0.016	0.533	**Heywood case**
taxesBH	68032.192	22231.070	0.002	Admissible

We notice that two residual variances are negative, which indicates a mathematical impossibility and signals that the maximum likelihood optimizer has converged to a point outside the parameter space. These are called Heywood cases, and they tend to arise when the likelihood surface is insufficiently constrained, which was truthfully predicted by Proposition 1 for near identified models in

 doi:10.20944/preprints202603.0278.v1

small samples. This point deserves closer attention, as a negative residual variance of -4350.076 is both a statistical red flag and a mathematical impossibility. In other words, any inference built on such a solution should not be treated as reliable regardless of what software reports as "converged".

Furthermore, Table 3 shows the parameter estimates. The standard errors obtained are large compared to the estimates, and no parameter achieves statistical frequentist significance.

Table 3. Frequentist ML Parameter Estimates.

Parameter	Estimate	Std.Err	z-value	Std.all
Measurement Model				
λ_1 (InstBH $\rightarrow$ ruleBH)	66.545	36.210	1.838	7.525
λ_2 (InstBH $\rightarrow$ GovEffBH)	0.010	0.006	1.570	0.063
λ_3 (ShadBH $\rightarrow$ GDP_BH)	0.153	0.062	2.474	1.245
λ_4 (ShadBH $\rightarrow$ taxesBH)	107.436	57.122	1.881	0.411
Structural Model				
β_1 (InstBH)	-0.055	0.048	-1.137	-0.050
β_2 (unemplBH)	0.314	0.235	1.332	0.558
β_3 (BroadMoneyGrowthBH)	0.007	0.006	1.243	0.133
β_4 (emplBH)	0.889	0.528	1.684	0.896

In addition, the standardized loading further reveals additional anomalies. For GDP_BH, $\lambda_{\text{std}} = 1.245$ exceeds unity, which is, safe to say, impossible under standard assumptions and results directly from the negative residual variance. Going further, we notice additional anomalies. For ruleBH, $\lambda_{\text{std}} = 7.525$ is implausible and reflects numerical instability. The key structural parameter, the effect of institutional quality on shadow economy, is estimated at $\beta_{\text{std}} = -0.050$ with $p = 0.255$, which finally indicates that there is no relationship whatsoever. Finally, these results exemplify the frequentist breakdown in small samples described in Section 1.2.1 and such a solution renders all inferential statements unreliable.

5.2. Bayesian Model Specification

We now estimate the same model using Bayesian methods implemented in `blavaan`.

5.2.1. Prior Specification

Based on economic theory, we specify:

- Factor loadings and regression coefficients: $\theta_i \sim \mathcal{N}(0, 10)$
- Residual standard deviations: $\sigma_i \sim \text{Gamma}(1, 0.5)$

With only 27 observations and 12 parameters, one finds oneself clearly in the aforementioned small sample regime. The $\mathcal{N}(0, 10)$ priors contribute minimal information relative to the likelihood but ensure that posterior remains proper in spite of near identification in certain directions.

5.2.2. Estimation and Convergence

Three different Bayesian SEM specifications were estimated using `blavaan` with the Stan backend. All models set `std.lv = TRUE` to fix latent variable variances to unity for identification. The specifications varied in MCMC sampling intensity and tuning parameters: (i) **Model 1** with 3 chains, 2,000 burn-in iterations and 10,000 posterior samples; (ii) **Model 2** with 9 chains, 7,000 burn-in iterations, 50,000 posterior samples, `adapt_delta = 0.99` and `max_treedepth = 25`, and (iii) **Model 3** with 3 chains, 4,000 burn-in iterations, 20,000 posterior samples, `adapt_delta = 0.99` and `max_treedepth = 15`.

Models 2 and 3 employed more aggressive tuning parameters (`adapt_delta = 0.99`) to handle the complex posterior geometry typical of small-sample SEM. Table 4 reports the convergence diagnostics.

 doi:10.20944/preprints202603.0278.v1

Table 4. Bayesian Model Convergence Diagnostics.

Diagnostic	Model 1	Model 2	Model 3
Chains	3	9	3
Burn-in	2,000	7,000	4,000
Posterior samples	10,000	50,000	20,000
Marginal Log-Lik.	−425.241	−173.405	−173.625
PPP	0.017	0.164	0.156
Max $\hat{R}$	1.000	1.067	1.051

Model 1, with minimal burn-in and fewer samples, shows poor convergence ($\hat{R} = 1.000$, but PPP $= 0.017$) and was estimated without the Stan target backend, which then resulted in worse marginal log-likelihood of −425.241. Models 2 and 3, however, employ Stan backend with more aggressive tuning (`adapt_delta = 0.99`). This yielded log-likelihoods which are comparable (−173.405 vs. −173.625) and PPP values (0.164 vs. 0.156). Hence, we select Model 2 as the benchmark for several reasons. First, it achieves marginally better marginal log-likelihood and a lower maximum $\hat{R}$ (1.067 vs. 1.051 for Model 3), and, in this way, providing a more conservative but better-sampled posterior. Truth be told, the additional computational cost is higher relative to Model 3, but justified by the use of 9 chains and 50,000 posterior samples. This helps us improve the exploration of the complex posterior geometry which is typical of small-sample SEMs.

Table 5 presents the key statistics across three specifications.

Table 5. Bayesian Measurement Model Estimates (Model 2).

Latent	Indicator	Est.	Post.SD	95% CI	Std.all
InstBH	ruleBH	0.847	0.297	[0.205, 1.378]	0.809
	GovEffBH	0.541	0.335	[−0.176, 1.176]	0.524
ShadBH	GDP_BH	0.247	0.238	[0.011, 0.836]	0.705
	taxesBH	0.132	0.154	[−0.027, 0.539]	0.378

Now all standardized loadings fall within plausible ranges, i.e. $0 < \lambda_{\text{std}} < 1$, which contrasts with the frequentist model anomalies. The 95% credible intervals (CI) for some loadings include zero, which can be justified given that this appropriately reflects estimation uncertainty rather than merely masking it behind inadmissible point estimates.

Table 6 reports the structural estimates.

Table 6. Bayesian Structural Model Estimates (Model 2).

Predictor	Est.	Post.SD	95% CI	Std.all
InstBH	−2.815	4.580	[−13.856, 6.129]	−0.942
unemplBH	−0.130	5.953	[−11.980, 11.929]	−0.043
BroadMoneyGrowthBH	−0.005	2.736	[−5.966, 6.001]	−0.002
emplBH	−0.161	9.021	[−17.354, 17.473]	−0.053

The institutional quality effect is successfully estimated at $\beta_{\text{std}} = -0.939$. This indicates a strong negative relationship with the shadow economy and the wide credible interval reflects genuine uncertainty given the constrained sample, but the posteriors concentrate on negative values, which is consistent with economic theory.

Table 7 presents the variances of the residuals. All estimates are positive and acceptable.

 doi:10.20944/preprints202603.0278.v1

Table 7. Bayesian Residual Variance Estimates (Model 2).

Variable	Est.	Post.SD	95% CI	Std.all
ruleBH	0.379	0.365	[0.001, 1.223]	0.346
GovEffBH	0.774	0.367	[0.021, 1.565]	0.726
GDP_BH	0.552	0.369	[0.007, 1.351]	0.504
taxesBH	0.936	0.324	[0.454, 1.712]	0.857

The residual variance of ShadBH (Std.all $= 0.112$) implies that the model explains approximately 88.8% ($(1 - 0.112) \cdot 100\%$) of the latent shadow economy construct.

6. Results and Discussion

6.1. Theoretical Implications

The BiH case study provides empirical validation for the theoretical framework developed in Sections 2-4. We summarize the key correspondences between theory and evidence:

1. **Asymptotic Equivalence under Ideal Conditions:** Theorems 3 establishes that Bayesian and frequentist methods converge to identical limiting distributions as $n \to \infty$. However, with $n = 27$ and $p = 16$, the BiH data are far from this asymptotic regime (n/p presenting the ratio between the number of observations and the number of parameters, which is used to assess the model fit in terms of overfitting on one side, and omitted variable bias on the other). The dramatic divergence between ML and Bayesian estimates ($\beta_{\text{std}} = -0.050$ vs. $\beta_{\text{std}} = -0.942$) illustrates that asymptotic equivalence provides no practical guidance in typical macroeconomic samples.
2. **Small-Sample Regime:** Theorem 4 predicts Bayesian dominance when $n \leq 5p$. Our application satisfies $n/p = 1.69 < 5$, placing it squarely in this regime. The theorem's prediction is confirmed: ML produces Heywood cases and meaningless estimates, while Bayesian estimation yields admissible, interpretable results.
3. **Identification and Regularization:** Proposition 1 states that near-singular Jacobians cause ML instability, while Bayesian priors maintain well-defined posteriors. The two negative variance estimates under ML (ruleBH: -4350.076; GDP_BH: -0.010) exemplify this instability. The same model estimated with $\mathcal{N}(0, 10)$ priors produces strictly positive variances throughout.
4. **Uncertainty Representation:** Theorem 8 predicts that Bayesian credible intervals achieve better finite-sample coverage than frequentist confidence intervals. While we cannot directly verify coverage without simulation, the frequentist intervals are constructed conditional on an inadmissible solution and are therefore meaningless. The Bayesian intervals, by contrast, reflect genuine posterior uncertainty over the admissible parameter space.

6.2. Substantive Findings

Beyond methodological validation, the analysis yields a substantive finding. Institutional quality shows a strong negative association with shadow economy activity in Bosnia and Herzegovina with $\beta_{\text{std}} = -0.942$. The macroeconomic control variables, unemployment $\beta_{\text{std}} = -0.043$), broad money growth ($\beta_{\text{std}} = -0.002$), and employment ($\beta_{\text{std}} = -0.053$) all show statistically non-credible effects conditional on institutional quality with 95% credible intervals spanning zero by a wide margin. This pattern is consistent with the small-sample setting and the dominant role of the institutional quality channel. This also might imply This might imply that policy interventions targeting institutional reform might be more effective than macroeconomic adjustments for reducing informality in the BiH context, which was substantially confirmed Medina and Schneider [3], and later on confirmed by recent cross-country analyses [25,26].

6.3. Practical Guidelines

Based on our theoretical and empirical findings, we offer the following practical recommendations. Bayesian SEM should be the default choice when $n/p < 10$, with weakly informative priors grounded

 doi:10.20944/preprints202603.0278.v1

in economic theory, and is particularly important for models with potential identification issues. Researchers should routinely report diagnostic checks: posterior predictive p-values (PPP around 0.5 indicates, though arbitrarily, acceptable fit), $\hat{R}$ statistics for all parameters, whether all variance estimates are positive, and whether all standardized loadings fall within $[0, 1]$. When large samples are available and $n/p > 20$ with well-identified models and no convergence issues, frequentist ML estimation remains a viable and computationally efficient choice, where asymptotic theory provides reliable inference.

6.4. Limitations

Several limitations require acknowledgment.First, the small sample size with $n = 27$ constrains statistical power in both Bayesian and frequentist frameworks. This thus limits generalizability beyond the context of Bosnia and Herzegovina, though our Bayesian approach clearly and explicitly addresses finite-sample inference. Second, nonlinearity is challenging to capture using this specification, meaning that structural and administrative breaks (common to transition and developing economies) may have been overlooked by the models, regardless of their specification. This could be resolved by enriching the model with additional explanatory, control, and dummy variables. However, doing so increases the risk of overfitting. This can potentially weaken the model performance significantly and even lead to non-convergence.

7. Conclusion

This paper attempts to mathematically analyze the Bayesian versus frequentist dilemma in structural equation modeling for economic applications. Hence, although both approaches converge asymptotically, fundamental differences emerge in finite samples. the finite sample, or more specifically, small sample, is characteristic of data related to economic activity. Let us present the main conclusions in that regard.

For large n, both methods yield equivalent results (Bernstein-von Mises theorem), but small samples exhibit certain differences. These are obvious in terms of bias, variance, and coverage properties. We will define this as **Convergence Dichotomy**.

Small sample settings, which are common in economics, provide a perfect playground for implicit regularization that frequentists simply cannot achieve. Bayesian priors act as implicit regularizes, providing superior performance in high-dimensional settings. This can be recognized as **Regularization Effect**.

Bayesian credible intervals provide an elegant way for **Uncertainty Quantification**. Frequentist confidence intervals can severely suffer from under-coverage in finite samples. On the flip side, Bayesian credible intervals maintain nominal coverage in finite samples.

The Bayesian approach provides **Identification Robustness**. Simply put, Bayesian framework assumes an elegant way of near-identification and weak identification. On the other hand, ML methods treat the whole issue sloppy. Readers could grasp this as a subjective remark. But, careful readers could find enough arguments.

In small samples, conventional frequentist fit indices can overstate adequacy. For example, we observe *a non-scaled* frequentist CFI of CFI ≈ 0.93 that appears "excellent," yet the same model estimated in BSEM yields posterior predictive p-values PPP $\in [0.02, 0.17]$ depending on sampling length and tuning. Hence, evidence of lackluster to, at best, modest fit. In other words, PPP is more sensitive to finite-sample and identification problems than CFI/TLI and should be prioritized when the effective sample size is limited. In our empirical application, we work with $n = 27$ observations and a moderately parameterized SEM; Bayesian re-estimation with longer chains/tighter `adapt` settings improves PPP from very low values (≈ 0.02) to still-cautious values (≈ 0.17), never matching the optimistic message conveyed by a high CFI.

Within the MIMIC framework, we have confirmed that the latent variable is inherently scale non-identified, even under correct specification and in the limit of an infinite sample size. Theorems on likelihood invariance and identification sets demonstrate that this indeterminacy is structural and

 doi:10.20944/preprints202603.0278.v1

cannot be resolved by asymptotic arguments alone. As a consequence, any empirical claim about latent levels necessarily relies on auxiliary identifying assumptions that lie outside the likelihood.

That said, if we zoom way out and only care about large samples, the divide between the frequentist and Bayesian approaches to SEM starts to disappear. The Bernstein–von Mises theorem establishes that, under regularity conditions and strong identification, Bayesian posteriors converge to the same Gaussian limit as ML estimators. Hence, in this sense, Bayesian and frequentist inference are asymptotically equivalent. However, the results developed in this paper make clear that asymptotics are precisely the regime in which macroeconomic MIMIC models rarely operate. Small samples, near-singular information matrices, and weak identification invalidate the quadratic approximations on which frequentist inference relies, while the likelihood remains flat. That being said, the key advantage of the Bayesian approach lies not in superior asymptotic efficiency but in its transparent treatment of identification and uncertainty in finite samples. In that sense, identification is achieved probabilistically rather than deterministically, and weakly identified parameters manifest as posterior dispersion rather than numerical instability, Heywood cases, or ill-conditioned Hessians.

Given the conclusions of this research and the constant dilemma between frequentist and Bayesian approaches, affected mostly by sample issues, be it the nominal or effective sample size, it is plausible to suggest that future research should focus on developing adaptive procedures that automatically select between Bayesian and frequentist approaches based on sample size, model complexity, and identification strength. Besides this, the development of computationally efficient variational Bayesian methods for large-scale economic structural equation modeling processes remains an open and important area for investigation.

Author Contributions: Conceptualization, Bojan Baškot; Methodology, Andrej Ševa and Vesna Lešević; Software, Bojan Baškot, Andrej Ševa and Bogdan Ubiparipović; Validation, Bojan Baškot, Andrej Ševa and Bogdan Ubiparipović; Formal analysis, Bojan Baškot, Andrej Ševa and Vesna Lešević; Investigation, Vesna Lešević; Data curation, Bogdan Ubiparipović; Writing – original draft, Bojan Baškot, Andrej Ševa, Vesna Lešević and Bogdan Ubiparipović; Writing – review and editing, Bojan Baškot; Supervision, Bojan Baškot. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jöreskog, K.G.; Goldberger, A.S. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association* **1975**, *70*, 631–639. https://doi.org/10.2307/2285946.
2. Schneider, F.; Buehn, A. Shadow Economy: Estimation Methods, Problems, Results and Open questions. *Open Economics* **2018**, *1*, 1 – 29.
3. Medina, L.; Schneider, F. Shadow Economies Around the World: What Did We Learn Over the Last 20 Years? IMF Working Paper WP/18/17, International Monetary Fund, 2018.
4. Young, G.A.; Smith, R.L. *Essentials of Statistical Inference*; Cambridge University Press, 2005.
5. Kleijn, B.J.K.; van der Vaart, A.W. The Bernstein–von Mises theorem under misspecification. *Electronic Journal of Statistics* **2012**, *6*, 354–381.
6. Gustafson, P. *Bayesian Inference for Partially Identified Models*; Chapman and Hall/CRC, 2015.
7. McNeish, D. On Using Bayesian Methods to Address Small Sample Problems. *Structural Equation Modeling: A Multidisciplinary Journal* **2016**, *23*, 750–773. https://doi.org/10.1080/10705511.2016.1186549.
8. Smid, S.C.; McNeish, D.; Miočević, M.; van de Schoot, R. Bayesian Versus Frequentist Estimation for Structural Equation Models in Small Sample Contexts: A Systematic Review. *Structural Equation Modeling: A Multidisciplinary Journal* **2020**, *27*, 131–161. https://doi.org/10.1080/10705511.2019.1577140.

 doi:10.20944/preprints202603.0278.v1

9. Jöreskog, K.G. *Structural analysis of covariance and correlation matrices*; Vol. 38, Springer, 1973; pp. 479–499. https://doi.org/10.1007/BF02293808.
10. Bollen, K.A. *Structural Equations with Latent Variables*; Wiley, 1989. https://doi.org/10.1002/9781118619179.ch1.
11. Breusch, T. Estimating the Underground Economy using MIMIC Models. Technical report, ANU, School of Economics, 2005. Revised November 2005. Working paper version widely circulated via EconWPA.
12. Breusch, T. Estimating the Underground Economy using MIMIC Models. *Unpublished manuscript* **2016**. Australian National University. Updated version available online.
13. Feige, E.L. Reflections on the Meaning and Measurement of Unobserved Economies: What do we really know about the "Shadow Economy"? *Journal of Tax Administration* **2016**, *2*, 5–41.
14. Kirchgässner, G. On Estimating the Size of the Shadow Economy. *German Economic Review* **2017**, *18*, 99–111. https://doi.org/10.1111/geer.12094.
15. Dybka, P.; Kowalczuk, M.; Olesiński, B.; Torój, A.; Rozkrut, M. Currency demand and MIMIC models: towards a structured hybrid method of measuring the shadow economy. *International Tax and Public Finance* **2019**, *26*, 4–40. https://doi.org/10.1007/s10797-018-9504-5.
16. Lee, S.Y. *Structural Equation Modeling: A Bayesian Approach*; Wiley, 2007. https://doi.org/10.1002/9780470024737.
17. Muthén, B.; Asparouhov, T. Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods* **2012**, *17*, 313–335. https://doi.org/10.1037/a0026802.
18. Asparouhov, T.; Muthén, B. Advances in Bayesian model fit evaluation for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal* **2021**, *28*, 1–14. https://doi.org/10.1080/10705511.2020.1764360.
19. Van Erp, S.; Mulder, J.; Oberski, D. Prior Sensitivity Analysis in Default Bayesian Structural Equation Modeling. *Psychological Methods* **2017**, *23*, 363–388.
20. Jacobucci, R.; Grimm, K. Comparison of Frequentist and Bayesian Regularization in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal* **2018**, *25*, 639–649.
21. Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*; Springer, 2005. https://doi.org/10.1007/978-3-540-27752-1.
22. Hox, J.; McNeish, D. Small samples in multilevel modeling. In *Small Sample Size Solutions: A Guide for Applied Researchers and Practitioners*; van de Schoot, R.; Miočević, M., Eds.; Routledge, 2020; pp. 215–225. https://doi.org/10.4324/9780429273872-18.
23. Hox, J.J.; van de Schoot, R.; Matthijsse, S. How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods* **2012**, *6*, 87–93. https://doi.org/10.18148/srm/2012.v6i2.5033.
24. Muthén, B.; Asparouhov, T. Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods* **2012**, *17*, 313–335.
25. Dell'Anno, R. Measuring the unobservable: Estimating informal economy by a structural equation modeling approach. *International Tax and Public Finance* **2023**, *30*, 247–277. https://doi.org/10.1007/s10797-021-09713-z.
26. Schneider, F. New COVID-related results for estimating the shadow economy in the global economy in 2021 and 2022. *International Economics and Economic Policy* **2022**, *19*, 299–318. https://doi.org/10.1007/s10368-022-00537-6.