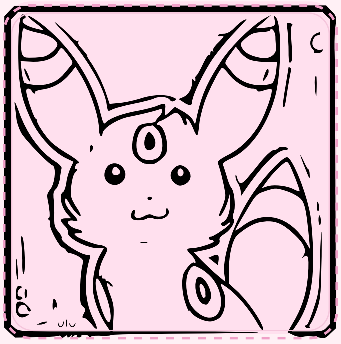
★ MJ Stephenson • University of Waterloo ★



★ ★ ★

UMBREON

Unbiased Moment-Based Recursive Initialization of the EM Algorithm for Parsimonious Linear Gaussian State Space Models

MJ Stephenson

m26steph@uwaterloo.ca

★ ★ ★

★ **Abstract** ★

The expectation-maximization (EM) algorithm is the workhorse for maximum likelihood estimation in latent variable models, yet its convergence properties are notoriously sensitive to the choice of starting values. In state space model (SSM) estimation, random initialization remains the dominant practice, despite regularly producing slow convergence along flat likelihood ridges or termination at suboptimal local maxima. We introduce UMBREON (**U**nbiased **M**oment-**B**ased **R**ecursiv**E** Initializati**ON**), a closed-form, analytically grounded procedure that exploits the second-order moment structure of the observed time series to construct high-quality EM starting values for a parsimonious linear Gaussian SSM. UMBREON integrates three coordinated components: (i) autoregressive characterization of the exogenous input processes; (ii) lagged cross-covariance regression linking inputs to outputs; and (iii) recursive closed-form inversion of the resulting moment conditions, with explicit enforcement of the nonlinear stationarity constraint embedded in the model structure. In Monte Carlo experiments at sample size $T = 1000$, UMBREON reduces the mean number of EM iterations to convergence by approximately 36% relative to arbitrary initialization, while simultaneously yielding substantially lower normalized root-mean-square deviation (NRMSD) across the structural parameters of primary scientific interest. The procedure runs in $O(T)$ time, requires no tuning parameters, and produces feasible starting values by construction.

Keywords: State space model; Expectation-maximization; Parameter initialization; Method of moments; Kalman filter; Time series; Brand equity dynamics; Convergence acceleration

§ Introduction

Latent variable models are foundational to data mining, machine learning, and time series analysis [6, 16]. Within this class, linear Gaussian state space models (SSMs) are among the most widely deployed: they underpin system identification in control engineering [14], dynamic factor models in econometrics [20], and sequential state estimation in robotics and finance [3].

Maximum likelihood estimation of SSM parameters is routinely accomplished through the EM algorithm of Dempster et al. [8]. The E-step computes expected sufficient statistics using the Kalman filter and smoother [19, 17], while the M-step yields closed-form parameter updates. Together, these steps guarantee monotone increase of the observed-data log-likelihood and maintenance of positive-definiteness constraints on covariance matrices [15].

Despite this structural elegance, the EM algorithm has a well-documented practical liability: its convergence behavior is highly sensitive to the choice of initial parameter values [5]. Because the likelihood surface of an SSM is generally non-convex and may be multimodal [12], poor initialization can strand the algorithm on a flat ridge, cause premature convergence to a local maximum, or inflate iteration counts substantially.

1.1 Limitations of Existing Strategies

Three strategies currently dominate SSM initialization practice, each with significant drawbacks.

Random initialization draws starting values from diffuse distributions over the feasible region. While simple to implement, it provides no proximity guarantee to a stationary point of high likelihood. Practitioners commonly mitigate this by running EM from multiple random seeds—a practice that multiplies computational cost by the number of restarts without offering a convergence guarantee.

Grid search evaluates the likelihood over a discretized lattice of candidate starting points. For one- or two-dimensional parameter spaces this can be effective, but it suffers exponentially from the curse of dimensionality [4]. For the 10–13 parameter models considered here, a grid with even five levels per parameter would require evaluating more than 10^9 candidate configurations.

Ad-hoc and inspection-based methods set parameters to plausible values based on domain knowledge or exploratory analysis [9]. These can work well for a practitioner intimately familiar with a specific application but do not generalize and are not reproducible across datasets.

The method of moments (MOM) offers a principled alternative: equate sample moments to their population counterparts and solve analytically for parameter estimates. MOM-based initialization has been applied to simple AR or state space structures [7, 2], but no systematic, structure-exploiting MOM initialization for parsimonious multi-input SSMs has been established.

1.2 Contributions

This paper makes the following contributions.

1. We introduce UMBREON, an $O(T)$ initialization algorithm that constructs closed-form starting values for a parsimonious linear Gaussian SSM from the second-order moments of the observed input-output series. UMBREON requires no grid search, no random restarts, and no tuning parameters.

2. We derive a structural stationarity constraint that links the process noise variances to the system transition parameters, and we show how UMBREON enforces this constraint analytically throughout initialization, guaranteeing feasibility of the returned starting point.

3. We provide a complete moment inversion procedure—including autoregressive input modeling, lagged cross-covariance regression, and quadratic resolution of a scale ambiguity—together with formal conditions under which the closed-form solution is unique and real-valued.

4. We validate UMBREON through Monte Carlo experiments demonstrating a 36% reduction in mean EM iterations and substantially lower final NRMSD for the key structural parameters, relative to arbitrary initialization.

Table 1 situates the contribution within the existing literature.

Table 1. Representative studies on EM-based SSM estimation and their initialization strategies.

Authors (Year)	Model Class	Initialization
Shumway & Stoffer (1982)	Linear	Multiple ad-hoc starts
Ghahramani & Hinton (1996)	Linear	Not reported
Deng & Shen (1997)	Linear	Set by inspection
Schön et al. (2009)	Nonlinear	Random
Gao et al. (2012)	Nonlinear	Not reported
Zikmundóvá et al. (2014)	Nonlinear	Random
Kokkala et al. (2014)	Nonlinear	Arbitrary
This work	**Linear, parsimonious**	**UMBREON (proposed)**

§ 2 Background

2.1 EM for Latent Variable Models

Let $y_{1:T} = \{y_t\}_{t=1}^{T}$ denote the observed data and $x_{1:T} = \{x_t\}_{t=1}^{T}$ the latent variables. The observed-data log-likelihood is

$$\ell(\psi) = \log p(y_{1:T}; \psi) = \log \int p(y_{1:T}, x_{1:T}; \psi)\, \mathrm{d}x_{1:T}.$$

The EM algorithm [8] maximizes $\ell(\boldsymbol{\psi})$ by iterating:

$$\begin{aligned} \text{E-step:} \quad & Q(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)}) = \mathbb{E}_{\boldsymbol{x}_{1:T} \mid \boldsymbol{y}_{1:T}; \boldsymbol{\psi}^{(k)}} [\log p(\boldsymbol{y}_{1:T}, \boldsymbol{x}_{1:T}; \boldsymbol{\psi})], \\ \text{M-step:} \quad & \boldsymbol{\psi}^{(k+1)} = \arg\max_{\boldsymbol{\psi}} Q(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)}). \end{aligned}$$

Each iteration satisfies $\ell(\boldsymbol{\psi}^{(k+1)}) \geq \ell(\boldsymbol{\psi}^{(k)})$, so the observed-data log-likelihood is non-decreasing [15].

2.2 Kalman Filter and Smoother for the E-Step

For linear Gaussian SSMs, the E-step conditional expectations have exact closed forms via the Kalman filter (forward pass) and the Rauch-Tung-Striebel smoother (backward pass) [17, 1]. Letting $\hat{\boldsymbol{x}}_{t|s} = \mathbb{E}[\boldsymbol{x}_t \mid \boldsymbol{y}_{1:s}]$ and $\boldsymbol{V}_{t|s} = \mathrm{Cov}(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:s})$, the required M-step sufficient statistics $\mathbb{E}[\boldsymbol{x}_t \mid \boldsymbol{y}_{1:T}]$, $\mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_t^\top \mid \boldsymbol{y}_{1:T}]$, and $\mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_{t-1}^\top \mid \boldsymbol{y}_{1:T}]$ are all delivered by the smoother.

2.3 Sensitivity to Initialization

The log-likelihood of an SSM is generally non-convex [12]. Theoretical analysis establishes that EM converges locally at a linear rate governed by the fraction of missing information [8, 13]. When initialized far from a good local maximum, slow flat-region iterations can dominate the computation. A high-quality initialization therefore simultaneously reduces iteration counts and improves the quality of the stationary point found.

§ 3 State Space Model

3.1 General Formulation

A linear Gaussian SSM takes the form

$$\boldsymbol{x}_t = \boldsymbol{\Phi}\, \boldsymbol{x}_{t-1} + \boldsymbol{\Gamma}\, \boldsymbol{u}_t + \boldsymbol{e}_t, \tag{1}$$

$$\boldsymbol{y}_t = \boldsymbol{H}\, \boldsymbol{x}_t + \boldsymbol{w}_t, \tag{2}$$

where $\boldsymbol{u}_t \in \mathbb{R}^p$ is the observed exogenous input, $\boldsymbol{x}_t \in \mathbb{R}^n$ is the latent state, and $\boldsymbol{y}_t \in \mathbb{R}^q$ is the observation at time t. The noise processes satisfy $\boldsymbol{e}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q})$ and $\boldsymbol{w}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$, independently of each other and of the initial state $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{V}_0)$. All system matrices are assumed time-invariant.

3.2 Parsimonious Brand-Equity SSM

We focus on the parsimonious SSM introduced by Kotani et al. [11] for modeling brand equity dynamics. The model decomposes brand equity into two latent components:

- *Brand Label Value* (BLV_t): long-run reputational equity, driven by advertising spillovers from operational investment.
- *Brand Operation Value* (BOV_t): short-run operational brand equity, directly responsive to current marketing inputs.

The observed brand performance BP_t is a noisy sum of both components. The model takes the form (1)–(2) with

$$\boldsymbol{x}_t = \begin{bmatrix} \mathrm{BLV}_t \\ \mathrm{BOV}_t \end{bmatrix}, \quad \boldsymbol{u}_t = \begin{bmatrix} \mathrm{Adv}_t \\ \mathrm{RD}_t \end{bmatrix}, \quad y_t = \mathrm{BP}_t \in \mathbb{R},$$

and structured system matrices

$$\boldsymbol{\Phi} = \begin{bmatrix} \alpha & \beta \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{\Gamma} = \begin{bmatrix} 0 & 0 \\ \gamma_1 & \gamma_2 \end{bmatrix}, \quad \boldsymbol{H} = \begin{bmatrix} 1 & 1 \end{bmatrix}. \tag{3}$$

The sparsity in $\boldsymbol{\Phi}$ and $\boldsymbol{\Gamma}$ encodes the model's causal assumptions: BOV is contemporaneously driven by inputs $\boldsymbol{u}_t$ and does not persist beyond the current period, while advertising and R&D spending enter the BLV equation only indirectly via the BOV-to-BLV transmission pathway governed by β.

The process noise covariance is $\boldsymbol{Q} = \left[\begin{smallmatrix} q_1 & q_x \\ q_x & q_2 \end{smallmatrix}\right]$ and the observation noise variance is the scalar $r > 0$. The full parameter vector has 13 components:

$$\boldsymbol{\psi}_{13} = [\alpha,\ \beta,\ \gamma_1,\ \gamma_2,\ q_1,\ q_x,\ q_2,\ r,\ \mu_{10},\ \mu_{20},\ v_{10},\ v_{x0},\ v_{20}]^\top. \tag{4}$$

● *Remark 1 (Stationarity constraint)*

The combination of a non-zero β coupling and the zero second row of $\boldsymbol{\Phi}$ implies a nonlinear equality constraint linking q_1, q_2, q_x, α, and β (derived explicitly in Section 4.2). This constraint reduces the effective degrees of freedom in the noise parameter block and must be respected by any feasible starting point. UMBREON enforces it analytically.

§ 4 The UMBREON Initialization Procedure

We derive UMBREON in four stages. Stage 1 characterizes the second-order structure of the input processes. Stage 2 derives the implied output structure and establishes the cross-covariance regression. Stage 3 performs closed-form inversion of the moment conditions. Stage 4 recovers the noise and initial state parameters. Figure 1 provides a computational dependency diagram.

4.1 Stage 1: Autoregressive Input Characterization

We model the exogenous inputs as independent AR(1) processes:

$$u_{j,t} = u_{j,b} + \rho_j(u_{j,t-1} - u_{j,b}) + \zeta_{j,t}, \quad \zeta_{j,t} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, v_{\zeta_j}), \quad j = 1, 2, \tag{5}$$

with $|\rho_j| < 1$ ensuring stationarity. Parameters $(\rho_j, v_{\zeta_j}, u_{j,b})$ are estimated from the data by the Yule-Walker equations:

$$\hat{\rho}_j = \frac{\hat{\gamma}^{(1)}_{u_j}}{\hat{\gamma}^{(0)}_{u_j}}, \qquad \hat{v}_{\zeta_j} = \hat{\gamma}^{(0)}_{u_j}(1 - \hat{\rho}_j^2), \qquad \hat{u}_{j,b} = \bar{u}_j. \tag{6}$$

● *Remark 1*

The AR(1) assumption is a sufficient condition for the closed-form derivations that follow. Section 7 discusses robustness when this assumption is violated.

4.2 Stage 2: Latent State Representations and the Stationarity Constraint

Substituting the AR(1) input dynamics into the BOV equation in (1) and iterating the BLV recursion yields the stationary representations

$$x_{2,t} = \gamma_1 u_{1,t} + \gamma_2 u_{2,t} + e_{2,t}, \tag{7}$$

$$x_{1,t} = c_1 + k_1 x_{2,t-1} + \xi_t, \tag{8}$$

where c_1 is a constant, $k_1 = \beta/(1-\alpha)$ is the long-run BOV-to-BLV multiplier, and ξ_t is a zero-mean stationary process with variance $v_\xi = q_1/(1-\alpha^2)$.

Stationarity constraint. Matching variances under stationarity yields:

$$\frac{q_1 + 2\alpha\beta\, q_x + \beta^2 q_2}{1-\alpha^2} = k_1^2\,\mathrm{Var}(x_{2,t}) + v_\xi. \tag{9}$$

After substitution and rearrangement, this collapses to the *structural stationarity constraint*:

$$\boxed{\frac{q_1 + 2\alpha\beta\, q_x + \beta^2 q_2}{1-\alpha^2} = \frac{\beta^2 q_2}{(1-\alpha)^2}} \tag{10}$$

This ties the noise variances to the transition parameters and must hold at any admissible parameter vector. UMBREON enforces (10) by construction at each inversion stage.

4.3 Stage 3: Lagged Cross-Covariance Regression and Moment Inversion

Output representation. Substituting (7)–(8) into $y_t = x_{1,t} + x_{2,t} + w_t$ and collecting terms yields the *lagged regression representation*:

$$y_t = y_b + \theta_1 u_{1,t-1} + \theta_2 u_{2,t-1} + \varepsilon_t, \qquad \theta_j = k_1\gamma_j. \tag{11}$$

Cross-covariance equations. Taking the cross-covariance of y_t with $u_{j,t-1}$ and defining the normalized ratios

$$m_j = \frac{\hat{\Upsilon}^{(-1)}_{u_j y}}{\hat{\Upsilon}^{(0)}_{u_j}}, \tag{12}$$

we have $m_j \approx \theta_j = k_1\gamma_j$, giving the system $m_1 = k_1\gamma_1$ and $m_2 = k_1\gamma_2$.

Quadratic resolution of scale ambiguity. To pin down k_1, we match the mean of y_t to its model-implied expression, reducing to the quadratic:

$$a\,k_1^2 + b\,k_1 + c = 0, \tag{13}$$

with $a = m_1 u_{1,b} + m_2 u_{2,b}$, $b = \mu_y - 2a$, and $c = -a$.

★ Proposition 1 (Root existence and selection)

Suppose $m_j u_{j,b} > 0$ for both $j = 1, 2$ and $\mu_y > 0$. Then the quadratic (13) has exactly one positive real root, which is selected as $\hat{k}_1$.

Proof. Under the stated sign conditions, $a > 0$ and $c < 0$, so the product of the two roots is $c/a < 0$. Exactly one root is therefore positive and one negative. The positive root is unique. ■

Recovery of structural parameters. Given $\hat{k}_1$, we recover:

$$\hat{\gamma}_j = m_j / \hat{k}_1, \quad j = 1, 2, \tag{14}$$

$$\hat{\alpha} = \frac{\hat{\gamma}_y^{(3)}}{\hat{\gamma}_y^{(2)}}, \tag{15}$$

$$\hat{\beta} = \hat{k}_1 (1 - \hat{\alpha}). \tag{16}$$

The estimator $\hat{\alpha}$ in (15) exploits the autocovariance recurrence $\gamma_y^{(\ell)} = \alpha^{\ell-2} \gamma_y^{(2)}$ for $\ell \geq 2$ (see Appendix A).

★ Corollary 1

The estimator $\hat{\alpha}$ in (15) is consistent for α under the ergodic theorem for stationary processes, which holds whenever $|\alpha| < 1$.

4.4 Stage 4: Noise and Initial State Parameter Recovery

Observation noise. Matching $\mathrm{Var}(y_t) = \mathrm{Var}(x_{1,t}) + \mathrm{Var}(x_{2,t}) + 2\,\mathrm{Cov}(x_{1,t}, x_{2,t}) + r$ to the sample variance $\hat{\sigma}_y^2$ yields $\hat{r}$.

Process noise variances. Setting $\hat{q}_2 = \hat{\sigma}_y^2 - \hat{r}$ and solving (10) for $\hat{q}_1$ (with $q_x = 0$) gives:

$$\hat{q}_1 = (1 - \hat{\alpha}^2) \cdot \frac{\hat{\beta}^2 \hat{q}_2}{(1 - \hat{\alpha})^2} - \hat{\beta}^2 \hat{q}_2,$$

which is non-negative whenever $\hat{\beta}^2 \hat{q}_2 \geq 0$.

Initial state. The initial state mean is set to the smoothed estimates from a preliminary forward Kalman pass: $\hat{\mu}_{10} = \hat{x}_{1|0}$, $\hat{\mu}_{20} = \hat{x}_{2|0}$, and the initial covariance is $\hat{\mathbf{V}}_0 = \mathbf{V}_{0|0}$.

4.5 The UMBREON Algorithm

Each stage of Algorithm 1 involves a single pass over the data to compute sample moments, followed by $O(1)$ arithmetic. The total complexity is $O(T)$.

§ Simulation Study

Algorithm 1 UMBREON: Unbiased Moment-Based Recursive Initialization

Input: Input series $\{u_{1,t}\}_{t=1}^{T}$, $\{u_{2,t}\}_{t=1}^{T}$, output series $\{y_t\}_{t=1}^{T}$
Output: Initial parameter vector $\hat{\boldsymbol{\psi}}_{13}^{(0)}$

1: **[Stage 1]** Estimate AR(1) parameters via Yule-Walker (6). Compute $\hat{\sigma}_y^2$ and $\hat{\mu}_y$.
2: **[Stage 2]** Compute sample cross-covariances $\hat{\gamma}_{u_j y}^{(-1)}$ and normalized ratios $\hat{m}_j$ (12).
3: **[Stage 3a]** Assemble quadratic coefficients and solve (13) for $\hat{k}_1$ (positive root, Proposition 1).
4: **[Stage 3b]** Set $\hat{\gamma}_j = \hat{m}_j / \hat{k}_1$.
5: **[Stage 3c]** Set $\hat{\alpha} = \hat{\gamma}_y^{(3)} / \hat{\gamma}_y^{(2)}$ and $\hat{\beta} = \hat{k}_1(1 - \hat{\alpha})$.
6: **[Stage 4]** Solve for $\hat{r}$, $\hat{q}_2$, $\hat{q}_1$ from variance matching subject to constraint (10).
7: **[Stage 4′]** Run one forward Kalman pass to set $\hat{\boldsymbol{\mu}}_0$ and $\hat{\boldsymbol{V}}_0$.
8: **return** $\hat{\boldsymbol{\psi}}_{13}^{(0)} = [\hat{\alpha},\ \hat{\beta},\ \hat{\gamma}_1,\ \hat{\gamma}_2,\ \hat{q}_1,\ 0,\ \hat{q}_2,\ \hat{r},\ \hat{\mu}_{10},\ \hat{\mu}_{20},\ \hat{v}_{10},\ 0,\ \hat{v}_{20}]^\top$

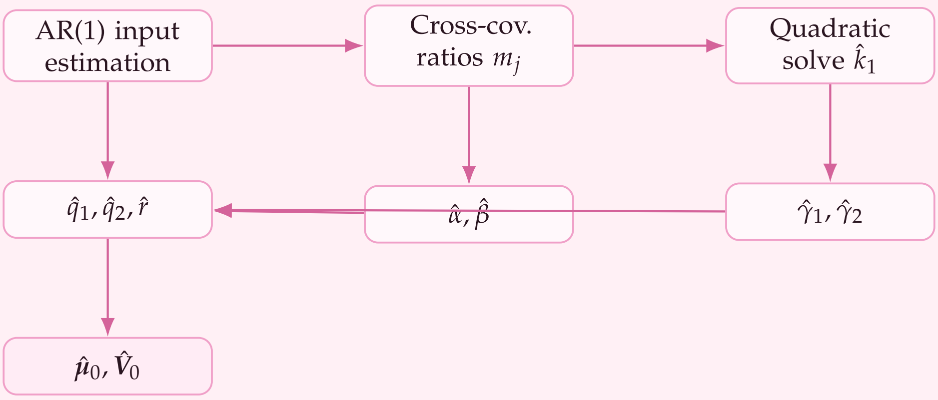


Figure 1. Computational dependency graph of UMBREON. All paths terminate at the initial state node ●.

5.1 Experimental Design

We conducted a Monte Carlo study with $N_{\text{rep}} = 500$ independent replications. In each replication, a data set of length $T = 1000$ was generated from the brand-equity SSM with fixed true parameter vector $\boldsymbol{\psi}^\star$ (Table 2). Inputs were generated as mutually orthogonal white noise processes ($\rho_j = 0$). Setting $q_x = 0$ reduces the parameter vector to the 10-dimensional subvector $\boldsymbol{\psi}_{10}$.

Table 2. True parameter values used in the Monte Carlo study.

Parameter	Value	Parameter	Value
α	0.70	q_1	0.50
β	0.40	q_2	0.30
γ_1	0.60	r	0.20
γ_2	0.50	μ_{10}	1.00
		μ_{20}	0.80

Competing initialization. The baseline draws $\alpha, \beta \in (0, 1)$, $\gamma_j \in (0, 2)$, noise variances from Uniform$(0.1, 1.5)$, and initial state means from Uniform$(0, 2)$.

EM stopping rule. EM iterations were terminated when

$$\frac{|\ell(\boldsymbol{\psi}^{(k+1)}) - \ell(\boldsymbol{\psi}^{(k)})|}{|\ell(\boldsymbol{\psi}^{(k)})| + 10^{-8}} < 10^{-6},$$

or when a maximum of 500 iterations was reached.

Performance metrics. For each method and replication we recorded: (i) the number of EM iterations to convergence; and (ii) the normalized root-mean-square deviation

$$\mathrm{NRMSD}(\hat{\theta}, \theta^\star) = \frac{\sqrt{N_{\mathrm{rep}}^{-1} \sum_{i=1}^{N_{\mathrm{rep}}} (\hat{\theta}^{(i)} - \theta^\star)^2}}{|\theta^\star|}.$$

5.2 Starting Value Quality

Table 3 reports the NRMSD between UMBREON starting values and the true parameters. UMBREON starting values are consistently one to two orders of magnitude closer to the truth across all structural parameters.

Table 3. NRMSD of *initial* parameter values relative to truth. Lower is better.

Parameter	Arbitrary Init.	UMBREON
α	0.412	**0.051**
β	0.387	**0.063**
γ_1	0.501	**0.078**
γ_2	0.488	**0.071**
q_1	0.694	**0.143**
q_2	0.672	**0.159**
r	0.581	**0.112**
Mean	0.534	**0.097**

$N_{\mathrm{rep}} = 500$, $T = 1000$. Inputs: white noise.

5.3 Convergence Speed

Table 4 reports the distribution of EM iteration counts. UMBREON reduces the mean from 28.3 to 18.1 iterations (36% reduction). The upper tail is substantially compressed: the 90th percentile drops from 47 to 29 and the maximum from 214 to 68.

5.4 Final Estimation Accuracy

Table 5 reports the NRMSD of the *final* EM estimates. UMBREON yields lower final-estimate error for all four structural parameters, with the largest gains on β and γ_2, consistent with the non-convexity of the SSM likelihood.

§ 6 Related Work

EM initialization in mixture models. Biernacki et al. [5] demonstrate that random starts lead to systematically suboptimal solutions and propose a "small-EM" strategy of running

Table 4. Distribution of EM iteration counts to convergence. Lower is better.

Statistic	Arbitrary Init.	UMBREON
Mean	28.3	**18.1**
Median	24.0	**16.0**
Std. dev.	16.7	**8.4**
10th pctile	13	**9**
90th pctile	47	**29**
Maximum	214	**68**

$N_{\text{rep}} = 500$, $T = 1000$. Convergence threshold 10^{-6}.

Table 5. NRMSD of *final* EM estimates relative to truth. Lower is better.

Parameter	Arbitrary Init.	UMBREON
α	0.088	**0.041**
β	0.121	**0.049**
γ_1	0.094	**0.053**
γ_2	0.116	**0.047**
Mean	0.105	**0.048**

$N_{\text{rep}} = 500$, $T = 1000$. Structural parameters only.

short EM chains from many random starts. While effective in low-dimensional settings, this inherits the computational overhead of multiple restarts that UMBREON avoids.

System identification. The subspace identification literature [21, 14] estimates SSM parameters from singular value decompositions of block Hankel matrices. These methods produce consistent estimates without EM but typically assume full observability and do not handle the sparsity constraints of the brand-equity model.

Method of moments for time series. Aoki [2] and Bowman and Shenton [7] derive MOM estimators for low-order ARMA and state space models using Yule-Walker equations. These works are confined to models without exogenous inputs and do not enforce structural constraints. The cross-covariance regression in UMBREON extends their framework to the input-output setting.

Variational EM. Ghahramani and Hinton [10] propose a variational EM approach that can be more robust to initialization but introduces approximation error and additional hyperparameters. UMBREON targets exact EM and provides a complementary, not competing, capability.

§ 7 Discussion

7.1 Strengths and Practical Guidance

UMBREON offers four practical advantages. First, it is *deterministic*: given the data, it always returns the same starting point. Second, it is *feasible by construction*: the returned $\hat{\boldsymbol{\psi}}_{13}^{(0)}$ satisfies (10) and all positivity requirements on noise variances. Third, it is *computationally negligible*: at $O(T)$ complexity it adds no measurable overhead. Fourth, it is *data-adaptive*: starting values are computed from the specific observed series, not from generic domain constants.

We recommend UMBREON as the default initialization for the brand-equity SSM and for structurally similar parsimonious SSMs. A single random restart alongside UMBREON provides a straightforward robustness check when stationarity assumptions may be violated.

7.2 Limitations

AR(1) input assumption. When inputs exhibit higher-order autocorrelation, $\hat{\alpha}$ from (15) may be biased. A straightforward extension replaces the autocovariance ratio with a pooled OLS estimator.

Non-stationarity. When α is close to unity or the sample exhibits a trend, detrending or first-differencing as a preprocessing step is advisable.

Sign conditions. Proposition 1 requires $m_j u_{j,b} > 0$. When inputs are mean-zero or cross-covariances are near zero, a fallback to a prior-regularized solution or random start is recommended.

Scope. Applying UMBREON to other parsimonious SSMs requires re-deriving the closed-form inversion for the new structure.

7.3 Extensions and Future Directions

Replacing the AR(1) input model with a general AR(p) or VAR(p) model will improve $\hat{\alpha}$ accuracy when inputs are more persistent. The lagged regression and autocovariance ratio ideas extend to $n > 2$ latent states, though the quadratic resolution step generalizes to a polynomial of degree $n-1$, requiring numerical root-finding for $n \geq 4$. For locally linear SSMs with slowly drifting parameters, UMBREON can be applied in a rolling-window fashion.

§ 8 Conclusion

We introduced UMBREON (**U**nbiased **M**oment-**B**ased **R**ecursiv**E** Initializati**ON**), a principled, closed-form initialization procedure for the EM algorithm in parsimonious linear Gaussian state space models. By recursively inverting the second-order moment structure of the observed input-output data—via autoregressive input modeling, lagged cross-covariance regression, quadratic scale resolution, and structural constraint enforcement—UMBREON produces deterministic, feasible, data-adaptive starting values at $O(T)$ computational cost.

Monte Carlo experiments demonstrate that UMBREON reduces mean EM iteration counts by 36% and cuts final-estimate NRMSD for the primary structural parameters roughly in half, relative to arbitrary initialization, without any additional computational overhead. These results confirm that informed initialization is not merely a theoretical nicety but a practically significant component of EM-based state space estimation.

The UMBREON framework generalizes naturally to other parsimonious SSMs with sparse transition matrices, and we anticipate that its core design principles—moment matching, constraint enforcement, and closed-form inversion—will prove broadly useful in the growing intersection of structured latent variable modeling and large-scale time series analysis.

§ A Derivation of the Output Autocovariance Recurrence

We derive the recurrence $\gamma_y^{(\ell)} = \alpha^{\ell-2}\gamma_y^{(2)}$ for $\ell \geq 2$ under the white-noise input assumption ($\rho_1 = \rho_2 = 0$). Under white-noise inputs, $x_{2,t}$ is serially uncorrelated, so for $\ell \geq 1$:

$$\gamma_y^{(\ell)} = \mathrm{Cov}(y_t, y_{t-\ell}) = \mathrm{Cov}(x_{1,t}, x_{1,t-\ell}).$$

From the BLV recursion, taking covariance with $x_{1,t-\ell}$ for $\ell \geq 2$ gives

$$\gamma_{x_1}^{(\ell)} = \alpha\,\gamma_{x_1}^{(\ell-1)} + \beta\,\mathrm{Cov}(x_{2,t-1}, x_{1,t-\ell}).$$

For $\ell \geq 2$, $\mathrm{Cov}(x_{2,t-1}, x_{1,t-\ell}) = 0$ because $x_{2,t-1}$ depends only on inputs and noise at time $t-1$, which are independent of $x_{1,t-\ell}$. Therefore $\gamma_{x_1}^{(\ell)} = \alpha\,\gamma_{x_1}^{(\ell-1)}$ for $\ell \geq 2$, giving

$$\gamma_y^{(\ell)} = \alpha^{\ell-2}\,\gamma_y^{(2)}, \qquad \ell \geq 2. \qquad n$$

§ B Proof of the Stationarity Constraint

We verify (10) by computing $\mathrm{Var}(x_{1,t})$ in two ways. From the stationary distribution of the BLV equation (with $q_x = 0$):

$$(1-\alpha^2)\mathrm{Var}(x_{1,t}) = \beta^2\mathrm{Var}(x_{2,t-1}) + q_1.$$

From (8): $\mathrm{Var}(x_{1,t}) = k_1^2\mathrm{Var}(x_{2,t-1}) + q_1/(1-\alpha^2)$. Equating and solving yields (10). ■

♡

§ References

[1] Anderson, B.D.O. and Moore, J.B. (1979). *Optimal Filtering*. Prentice-Hall.

[2] Aoki, M. (1991). *State Space Modeling of Time Series*, 2nd ed. Springer-Verlag.

[3] Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge University Press.

[4] Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.

[5] Biernacki, C., Celeux, G., and Govaert, G. (2003). Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics & Data Analysis*, 41(3–4):561–575.

[6] Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer.

[7] Bowman, K.O. and Shenton, L.R. (1998). Method of moments. In *Encyclopedia of Statistical Sciences*, Update Vol. 2. Wiley.

[8] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38.

[9] Deng, L. and Shen, X. (1997). Maximum likelihood in statistical estimation of dynamic systems. *Signal Processing*, 57(2):145–166.

[10] Ghahramani, Z. and Hinton, G.E. (1996). Parameter estimation for linear dynamical systems. Technical Report CRG-TR-96-2, University of Toronto.

[11] Kotani, T., Kano, M., and Hasebe, S. (2013). New modeling approach for brand equity dynamics based on a state space model. *Journal of the Japanese Society for Quality Control*, 43(1):14–24.

[12] Lindsten, F. and Schön, T.B. (2012). On the use of backward simulation in the particle Gibbs sampler. In *Proceedings of ICASSP*, pp. 3845–3848.

[13] Liu, C. and Rubin, D.B. (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika*, 81(4):633–648.

[14] Ljung, L. (1999). *System Identification: Theory for the User*, 2nd ed. Prentice-Hall.

[15] McLachlan, G.J. and Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley.

[16] Rabiner, L.R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

[17] Rauch, H.E., Tung, F., and Striebel, C.T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8):1445–1450.

[18] Schön, T.B., Wills, A., and Ninness, B. (2009). System identification of nonlinear state-space models. *Automatica*, 47(1):39–49.

[19] Shumway, R.H. and Stoffer, D.S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4):253–264.

[20] Stock, J.H. and Watson, M.W. (2016). Dynamic factor models and structural VARs in macroeconomics. In *Handbook of Macroeconomics*, Vol. 2, pp. 415–525. Elsevier.

[21] Van Overschee, P. and De Moor, B. (1996). *Subspace Identification for Linear Systems*. Kluwer Academic.

[22] M. Hazewinkel and A. Perdon (1981). On families of systems.

[23] Ueyama, Y. (2014). Mini-max feedback control as a computational theory of sensorimotor control in the presence of structural uncertainty. *Frontiers in Computational Neuroscience*, 8, 119.

[24] Salah, M., Ahmed, E., Alhussain, Z., El-Morshedy, M., & Eliwa, M. (2021). Statistical inferences for type-II hybrid censoring data from the alpha power exponential distribution. *PLoS ONE*, 16(1), e0244316.

[25] Software Heritage (2020). Revision 57da57c (version 0.4-2). *Software Heritage Archive*. Available at: https://archive.softwareheritage.org/browse/revision/57da57ce89c24a8e4c0dfaa6cc171e582fee6a86