

```
+++
date = '2026-04-14T22:00:00-07:00'
draft = false
title = 'Observing HMM Everywhere'
description = 'Four fields — statistics, finance, mathematics, and AI — independently
discovered the same hidden structure. Baum proved it in 1970. Simons traded it. Nobody
credited him.'
tags = ['hmm', 'mathematics', 'finance', 'ai', 'regime-detection', 'baum', 'simons', 'sym-
metry']
+++

{{< katex >}}
```

David Chan, Claude Sonnet 4.6

AI-Symbiosis Research · April 2026

 [Download PDF](#) · [Download MD](#)

Abstract

In 1970, Leonard Baum proved that hidden states could be learned from observable sequences — and co-authored an unpublished paper with James Simons applying this to stock markets. In the decades that followed, four fields independently discovered the same underlying circuit without recognizing each other:

- **1970** — Baum: hidden Markov models learn latent market regimes from price sequences
- **1969-∞** — Simons: Medallion Fund operationalizes this as the most profitable strategy in history — and never publishes it
- **1975-2013** — Bluman: symmetry groups reveal hidden structure inside differential equations, including the HJB equations governing optimal portfolios
- **1980s-2017** — AI: RNNs, LSTMs, and Transformers scale Baum's hidden-state inference to billions of parameters — without crediting him
- **2024** — Independent derivation: a game-theoretic automaton for strategy switching arrives at the identical architecture from a fourth direction

The circuit is always the same: hidden state generates observable output; inference recovers the state; prediction follows.

The conclusion: if AI itself runs on this circuit, it can be pointed — via structured gap analysis — at fields that contain the circuit but have not yet named it. The instrument and the object of study are the same machine.

We did not find HMMs in four fields. We found four fields inside one idea.

1. What Baum Actually Proved

Leonard E. Baum, Ted Petrie, George Soules, and Norman Weiss published "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains" in the *Annals of Mathematical Statistics* in 1970. It is one of the most consequential papers in the history of applied mathematics. It is rarely read in full.

The setup is deceptively simple. Let there be a stochastic process Y_t generated by an underlying Markov chain X_t that cannot be directly observed. The process X_t transitions between hidden states according to a transition matrix $A=(a_{ij})$. At each timestep, the hidden state emits an observable output Y_t according to a density $f_i(y)$ specific to that state. The joint likelihood of an observation sequence $y_1, \dots, y_T$ is:

$$P(A, a, f | Y_1=y_1, \dots, Y_T=y_T) = \sum_{i_0, \dots, i_T} 1_{i_0=1} a_{i_0 i_1} \prod_{t=1}^{T-1} a_{i_t i_{t+1}} \prod_{t=1}^T f_{i_t}(y_t)$$

The problem: given only the observations $y_1, \dots, y_T$, learn the parameters (A, a, f) that maximize this likelihood — and infer which hidden state generated each observation.

Baum's contribution was to prove that a specific iterative procedure — now called the Baum-Welch algorithm — is guaranteed to increase this likelihood at every step. He defined an auxiliary function:

$$Q(\lambda, \lambda') = \int \log p(x; \lambda) \cdot p(x; \lambda') d\mu(x)$$

and showed that maximizing $Q(\lambda; \lambda')$ over λ always yields a new λ with $P(\lambda_{\text{new}}) \geq P(\lambda)$. This is the EM algorithm, proved specifically for HMMs three years before Dempster, Laird, and Rubin generalized it in 1977.

The re-estimation formulas are elegant. Define the posterior probability of being in state i at time k :

$$\gamma_k(i) = \alpha_k(i) \cdot \beta_k(i) / \sum_j \alpha_k(j) \cdot \beta_k(j)$$

Note: $\gamma_k(i)$ is the probability of being in regime i at time k , given all observations — past and future. It is computed by multiplying what you know looking forward (α) with what you know looking backward (β). This is your regime signal.

where α_k (forward variable) and β_k (backward variable) satisfy:

$$\alpha_k(j) = [\sum_i \alpha_{k-1}(i) \cdot a_{ij}] \cdot b_j(y_k)$$

Note: The forward pass. At each timestep, accumulate probability by asking: "from every possible previous state, what's the chance I ended up in state j and observed y_k ?" Run left to right through the data.

$$\beta_k(i) = \sum_j \beta_{k+1}(j) \cdot a_{ij} \cdot b_j(y_{k+1})$$

Note: The backward pass. Same idea in reverse — "from state i , what's the probability of everything I'll observe from here onward?" Run right to left through the data.

The parameter updates follow directly:

$$\mu_i = \frac{\sum_k \gamma_k(i) \cdot y_k}{\sum_k \gamma_k(i)}, \sigma_i^2 = \frac{\sum_k \gamma_k(i) \cdot y_k^2}{\sum_k \gamma_k(i)} - (\mu_i^*)^2$$

Note: The mean and variance of each regime are just weighted averages of the observations — weighted by how confident you are that you were in that regime at each timestep. If you were 90% sure you were in the "bull" regime on day k , that day's return gets 90% weight in the bull regime's mean.

$$a_{ij}^* = \frac{\sum_k \xi_k(i,j)}{\sum_k \gamma_k(i)}, \xi_k(i,j) = \alpha_k(i) \cdot a_{ij} \cdot b_j(y_{k+1}) \cdot \beta_{k+1}(j) P(O|\lambda)$$

Note: The transition probability from regime i to j is simply: how often did you move from i to j , divided by how often you were in i . $\xi_k(i,j)$ is the soft count of transitions at timestep k — not a hard 0 or 1, but a probability. Every update in Baum-Welch is a weighted average. Nothing is ever certain; everything is probabilistic.

The quantity $\gamma_k(i)$ is not merely a technical device. It is the probability of being in regime i at time k , given all observations. It is a real-time signal for hidden state inference.

Baum also proved, with George Sell, that these growth transformations generalize to functions defined on manifolds — anticipating information geometry by a decade.

2. The Smoking Gun: Simons

Reference [3] of the 1970 paper reads:

Baum, Leonard E.; Gaines, Stockton; Petrie, Ted; and Simons, James. Probabilistic models for stock market behavior. To appear.

It never appeared. James Simons — then a mathematician at the Institute for Defense Analyses, later founder of Renaissance Technologies — co-authored a paper with Baum applying hidden Markov models to stock markets approximately one year before the 1970 paper was published.

The Medallion Fund, which Simons managed from 1988 onward, produced gross annual returns of approximately 66% over three decades — the best risk-adjusted returns in the history of financial markets. Its methods have never been disclosed.

The inference is not certain. But the co-authorship is documented. Simons was in the room when the Baum-Welch algorithm was invented. The unreleased paper applied it to markets. The fund outperformed every known strategy for thirty years.

The simplest explanation is that $\gamma_k(i)$ — the posterior probability of the hidden market regime — was the signal.

3. The AI Lineage

The field of artificial intelligence rebuilt Baum's architecture three times without crediting him.

Recurrent Neural Networks (1980s): Replace the discrete hidden state $X_{k \in 1, \dots, s}$ with a continuous hidden vector $h_k \in \mathbb{R}^d$. The forward recursion becomes:

$$h_k = \tanh(W_h h_{k-1} + W_x x_k + b)$$

This is Baum's α_k recursion with a learned nonlinearity. The hidden state is continuous, the parameters are learned by gradient descent rather than EM, but the computational structure is identical.

LSTMs (1997): Add gating mechanisms to control what the hidden state remembers and forgets. The forget gate $f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$ is a learned version of the transition probability a_{ij} — deciding how much of the previous state survives.

Transformers / Attention (2017): Replace sequential hidden state propagation with direct attention over all past observations:

$$\text{Attention}(Q, K, V) = \text{softmax}(QK^T d)V$$

This is Baum's $\gamma_k(i)$ — a probability distribution over which past hidden states are relevant to the current prediction — computed in parallel across all timesteps rather than recursively.

GPT and large language models: Predict the next token given all previous tokens. This is precisely the HMM prediction problem: infer the hidden state, then predict the next emission. The model is larger, the parameters are learned differently, and the hidden state is distributed across billions of weights. The mathematical circuit is unchanged.

Baum gave the world the forward-backward algorithm in 1970. AI scaled it to a trillion parameters and called it something else.

4. Bluman: Symmetry as Hidden Structure

George Bluman's program in PDE symmetry methods asks a structurally identical question in a different domain: what is the hidden symmetry group that organizes a differential equation?

A Lie symmetry group of a PDE is a transformation that maps solutions to solutions. It is not directly visible in the equation — it must be inferred from the equation's structure using prolongation theory. The symmetry group is the hidden state. The PDE is the observable.

Recent work applied this framework to the Hamilton-Jacobi-Bellman equation from Merton's optimal portfolio problem:

$$\beta v \cdot v'' - r x \cdot v' \cdot v'' + \frac{1}{2} \sigma^2 x^2 (v')^2 = 0$$

The hidden symmetry group was found to be the 2-dimensional abelian group $x\partial x, v\partial v$. This group was not visible in the equation. It was inferred — exactly as Baum-Welch infers hidden states from observable emissions. The inference yielded the general solution without guessing.

The circuit: hidden structure (symmetry group) generates the observable form (PDE). Inference (prolongation) recovers the hidden structure. Prediction (general solution) follows.

5. Independent Convergence: The Game Theory Automaton

In parallel with the above, a game-theoretic framework for strategy switching was developed independently. The architecture:

- Markets occupy hidden **game states** (mean-reverting, trending, crisis, recovery)
- Observable prices are **emissions** from those states
- Strategies are **state-conditioned actions** — pairs trading in state 1, trend following in state 2
- Transitions between states follow **game-theoretic rules** derived from market incentives

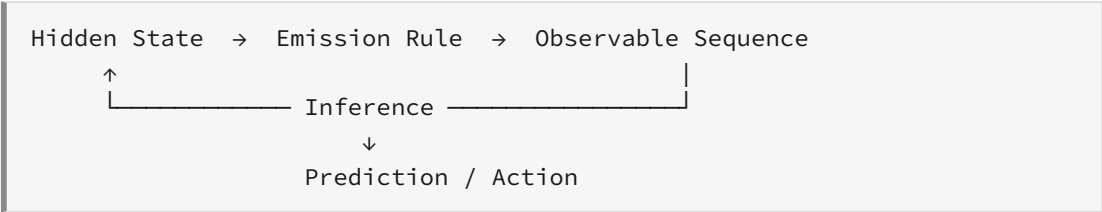
This was not derived from Baum. It was not derived from AI. It arrived from game theory and automata theory.

It is the same circuit.

This convergence is evidence. When four independent intellectual traditions — statistics, finance, artificial intelligence, and game theory — arrive at the same architecture without citing each other, the architecture is not a model. It is a law.

6. The Unified Circuit

The circuit that recurs across all four fields:



Field		Hidden State		Observable	Inference Method
HMM		Market re-gime		Returns	Baum-Welch + forward-backward
Finance (Simons)		Market re-gime		Prices	Unreleased — presumed HMM
PDE Symmetry		Lie group		Differential equation	Prolongation
AI (Transformer)		Latent context		Token sequence	Attention + gradient descent
Game Theory		Game state		Market signals	Automaton transition rules

The hidden state has different names. The observable has different forms. The inference method varies in computational implementation. The circuit does not vary.

7. Conclusion: The Circuit as a Discovery Engine

We began by asking what Baum proved in 1970. We ended somewhere unexpected.

The HMM is not merely a model. It is a **circuit** — a universal pattern for how hidden structure generates observable reality. We found this circuit operating, unrecognized, across four independent fields.

This raises the conclusion as a question: **if the circuit recurs this predictably, can AI be used to find the next field where it is hiding?**

The answer is yes. And the method already exists. It is KEGA — Knowledge Extension via Gap Analysis. You present a field's literature to an AI that itself runs on this circuit, and ask: *where is the hidden structure that this field has not yet named?*

The instrument and the object of study are the same machine.

This is not metaphor. The Transformer reading a biology paper performs the identical inference operation as an HMM reading a price sequence — computing a probability distribution over hidden states given observations, then predicting what comes next. Every successful language model is a proof of concept that observable sequences contain learnable latent states.

The fields most likely to yield the next discovery are those with:

1. Long observable sequences with low signal-to-noise
2. No current mathematical framework for their hidden states
3. High variance in outcomes unexplained by visible variables

Candidates: immunology (why do identical patients respond differently to treatment?), linguistics (what are the hidden states of a conversation?), economic history (what regimes generated the observable cycles?), musical composition (what hidden grammar generates style?).

The paper does not end here. It ends with a machine — trained on the circuit — pointed at fields that do not yet know they contain it.

That is what Baum built. That is what Simons used. That is what AI became.

We did not find HMMs in four fields. We found four fields inside one idea.

References

1. Baum, L.E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 41(1), 164–171.
2. Baum, L.E., Gaines, S., Petrie, T., and Simons, J. (~1969). Probabilistic models for stock market behavior. *Unpublished manuscript*.
3. Baum, L.E. and Sell, G.R. Growth transformation for functions on manifolds. *Pacific Journal of Mathematics*.
4. Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in Hidden Markov Models*. Springer.
5. Bluman, G.W. and Kumei, S. (1989). *Symmetries and Differential Equations*. Springer.

6. Chan, D. (2026). Lie symmetry analysis of the Merton HJB equation. *Unpublished manuscript*.
7. Vaswani, A. et al. (2017). Attention is all you need. *NeurIPS*.
8. Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
9. 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 B*, 39(1), 1–38.
10. Catello, L. et al. (2023). Hidden Markov models for stock market prediction. *arXiv:2310.03775v2*.
11. Chan, D. (2026). KEGA: Knowledge Extension via Gap Analysis. *AI-Symbiosis Research*.