Smarter Than 60/40:

Building a Regime-Switching Portfolio with Machine Learning

A 10-year backtest using Canadian ETFs, Hidden Markov Models, and walk-forward validation

The Problem with Static Allocation

The 60/40 portfolio: *60% equities, 40% bonds* has been the default recommendation for balanced investors for decades. It is simple, cheap to implement, and has a long track record. But it has one fundamental flaw: the market does not behave the same way all the time, and a portfolio that ignores that is leaving risk-adjusted return on the table.

Consider the difference between 2017 and 2022. In 2017, equities ground steadily upward with almost no volatility, a year where holding 40% in bonds was a drag. In 2022, both equities and bonds fell simultaneously in response to aggressive rate hikes, resulting in a year where a conventional 60/40 investor had nowhere to hide and suffered one of the worst drawdowns in the strategy's history.

> **Core insight:** If we can identify which market environment we are currently in, we can adjust our portfolio accordingly — holding more risk assets when conditions favour them, and rotating defensively before volatility arrives rather than after.

This post walks through how I built a regime-switching allocation strategy for a Canadian investor using a Hidden Markov Model, mean-variance optimisation, and rigorous out-of-sample testing. The results are promising and I will also be honest about the limitations.

What Is a Market Regime?

A market regime is a persistent state of financial market conditions. Think of it as the current backdrop against which all asset prices are moving. In practice, markets tend to cluster into recognisable environments: calm trending periods, choppy uncertain periods, and acute stress events.

The key word is **persistent**. Regimes are not random daily noise. A bull market does not last one day and then flip to a crisis. It sustains itself for weeks or months due to underlying economic conditions. This persistence is exactly what makes regime-switching models useful for portfolio management.

For this project I defined three regimes:

- **Low-Vol Bull:** Realised equity volatility is low, momentum is positive, credit spreads are tight. This is the "risk-on" environment where equities tend to outperform.
- **Medium-Vol Normal:** Moderate volatility, mixed signals. The default state which is neither euphoric nor panicked.
- **High-Vol Crisis:** Volatility spikes, correlations converge toward 1, credit spreads widen sharply. Capital preservation takes priority.

Across the full 2015–2024 dataset, the model assigned 52.3% of trading days to the Bull regime, 45.6% to Normal, and 2.1% (52 days) to Crisis, a distribution that feels intuitively right for a decade that contained mostly calm markets punctuated by two sharp dislocations.

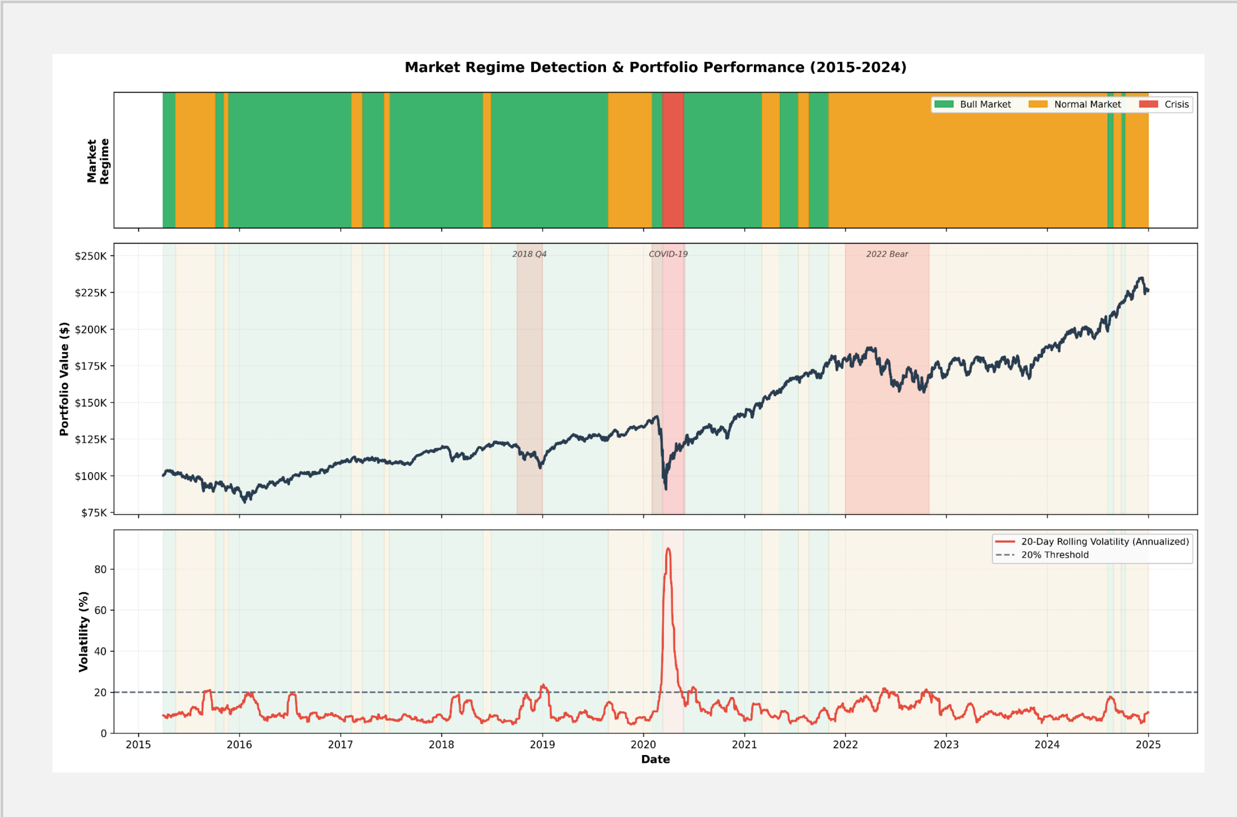

Figure 1: Detected market regimes across the 2015–2024 period. Green = Bull, Orange = Normal, Red = Crisis.

The Model: Hidden Markov Model

To detect which regime the market is currently in, I used a **Gaussian Hidden Markov Model (HMM)**. HMMs are a statistical framework that assumes observed market data is generated by an unobserved (hidden) underlying state, and that those states transition from one to another according to fixed probabilities.

The intuition is straightforward: we cannot directly observe whether the market is in a bull or crisis regime, but we can observe signals that are characteristic of each. The HMM learns these signal patterns from historical data and then assigns probabilistic regime labels to each day going forward.

Features fed to the model

Rather than feeding raw returns into the model, I engineered ten features designed to capture different dimensions of market stress:

- **Realised volatility (20-day rolling)** for equities, bonds, and their correlation: the most direct measure of current turbulence.
- **60-day momentum** for equities, bonds, and gold: captures the directional trend of each asset class.
- **Credit spread proxy** (HYG/LQD volatility ratio): a forward-looking stress indicator. When high-yield bonds become disproportionately more volatile than investment-grade, it signals rising credit risk and typically precedes broader equity drawdowns.
- **Yield curve slope** (10-year Treasury minus 3-month T-bill): an inverted yield curve has preceded every US recession in the modern era. Adding this gives the model a macro anticipatory signal rather than purely backward-looking inputs.

Posterior probability blending and the dwell filter

Rather than hard-switching between regimes based on the model's single best guess, the strategy uses **posterior probability blending**. If the model is 70% confident we are in a bull regime and 30% confident we are in a normal regime, the portfolio holds 70% of the bull-regime weights blended with 30% of the normal-regime weights. This produces smoother transitions and fewer unnecessary trades.

A **minimum dwell filter** adds further stability: the model must predict a new regime for at least five consecutive days before accepting the switch. In the full in-sample prediction, this suppressed 92 premature single-day regime flips though the precise count varies across walk-forward splits since each re-trains independently.

Portfolio Construction: The Five ETFs

The strategy is built entirely from five Toronto Stock Exchange-listed ETFs, making it directly actionable for a Canadian investor with a standard brokerage account and no currency conversion required:

- **XIU.TO:** iShares S&P/TSX 60. Canadian equity exposure to the 60 largest companies on the TSX.
- **VFV.TO:** Vanguard S&P 500 Index. US equity exposure in CAD-hedged form.
- **XEF.TO:** iShares MSCI EAFE. International developed market exposure (Europe, Australasia, Far East).
- **XBB.TO:** iShares Canadian Bond Index. Broad Canadian investment-grade bond exposure.
- **CGL-C.TO:** iShares Gold Bullion (CAD-hedged). Gold as a crisis hedge and inflation diversifier.

How weights are determined for each regime

For the Bull and Normal regimes, portfolio weights are derived using **mean-variance optimisation (MVO)**. MVO is the classic framework for finding the portfolio that maximises expected return for a given level of risk. However, naive MVO has a well-documented problem: it is extremely sensitive to the expected return inputs, which are estimated from noisy historical data and tend to produce unstable, concentrated portfolios.

To address this I applied two corrections. First, Ledoit-Wolf shrinkage on the covariance matrix, which stabilises the correlation structure between assets. Second, mean return shrinkage, which involves pulling each asset's expected return toward the average across all assets, reducing the optimiser's overconfidence in any single asset's historical edge. The intensity of shrinkage scales with regime data scarcity: 20% for the Bull regime (plenty of data), 30% for Normal, and 70% for Crisis.

For the **Crisis regime** I made a deliberate choice to hard-code a fixed defensive allocation rather than optimise it. The full ten-year dataset contains only approximately 52 crisis days — statistically insufficient to estimate a reliable covariance matrix for five assets. The fixed allocation is **55% bonds, 25% gold, and 20% equities**.

Gold is capped at 25% across all regimes. The 2015–2024 period was exceptionally favourable for gold, and allowing the optimiser unconstrained access to gold allocation would make the strategy's performance overly sensitive to that single asset's continued outperformance. This is something I would not want to bet my own money on.

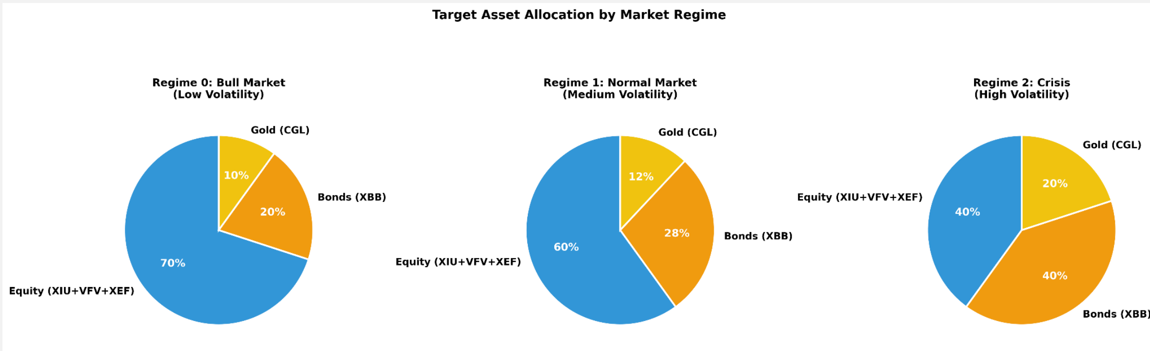

Figure 2: Target allocations by regime. Note the deliberate gold cap at 25% even in defensive regimes, and the hard-coded crisis allocation.

Results

Full-sample performance (2015–2024)

Across the full 10-year period, the optimised HMM strategy grew $100,000 to approximately **$284,000**, compared to $186,000 for a 60/40 benchmark with annual rebalancing (a total return of 184% vs 86%). The strategy delivered an annualised return of **11.4%** against the benchmark's 6.6%. Volatility was modestly higher 9.7% annualised versus 9.1% for the benchmark reflecting the strategy's greater equity exposure during bull regimes. The risk-adjusted improvement is therefore real but not free: the higher Sharpe ratio (0.97 vs 0.55) comes alongside somewhat larger absolute swings.

The Sortino ratio, which penalises only downside volatility, was **1.23 versus 0.58**. The strategy's worst drawdown was **-17.8%** versus -20.9% for the 60/40 benchmark.

Figure 3: Cumulative portfolio value from $100,000 starting capital. Shaded regions mark the COVID crash (2020), 2018 Q4 correction, and 2022 rate shock.

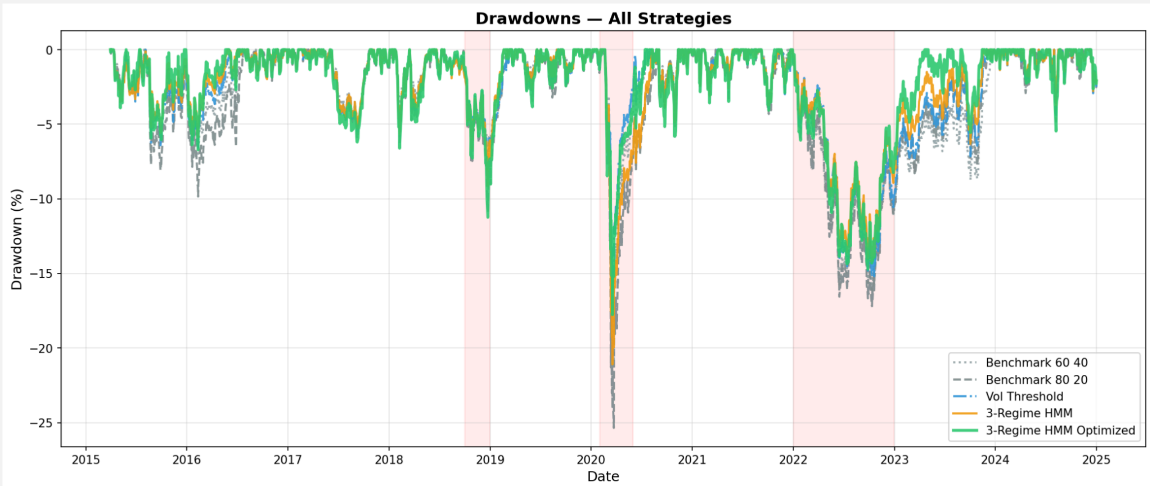

Figure 4: Portfolio drawdown over time. The strategy's worst drawdown was -17.8% versus -20.9% for the 60/40 benchmark.

Walk-forward out-of-sample validation

Full-sample backtests are susceptible to overfitting. The more important test is walk-forward analysis: train the model on data up to a cutoff date, apply it to the subsequent year, then roll forward and repeat. This simulates what an investor would have actually experienced, with no knowledge of the future.

The strategy was evaluated across four non-overlapping one-year test periods between 2020 and 2024, covering the COVID recovery, the post-COVID bull market, the 2022 rate shock, and the 2023–2024 recovery:

Strategy	Sharpe	Ann. Return	Ann. Vol	Max DD	Calmar
60/40 Benchmark	0.88	9.5%	8.3%	-7.3%	2.00
80/20 Benchmark	1.04	13.1%	10.3%	-8.2%	2.30
Vol Threshold	1.08	11.1%	8.8%	-7.3%	2.17
3-Regime HMM	1.06	11.0%	8.5%	-6.8%	2.22
3-Regime HMM Optimized	**1.27**	**15.0%**	**10.6%**	**-8.3%**	**2.79**

All metrics are averages across the four walk-forward test periods. Sharpe ratios use a constant 2% risk-free rate (see note below).

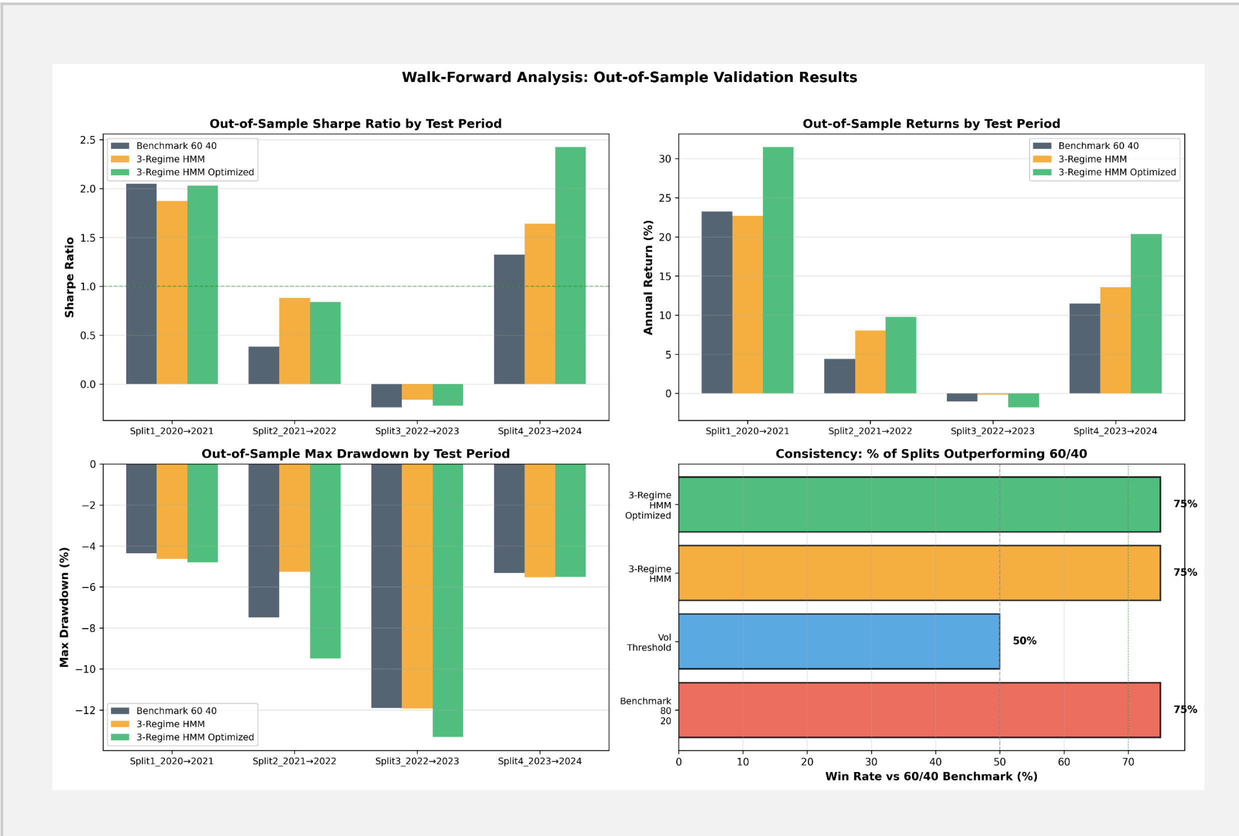

Figure 5: Walk-forward Sharpe ratios across four test periods. Each bar represents one year of out-of-sample performance.

The optimised strategy outperformed the 60/40 benchmark in ***three out of four*** out-of-sample periods, with an average excess Sharpe of +0.39. This is the result I trust, not the full-sample backtest number.

> **Note on Sharpe rate consistency**
>
> Walk-forward Sharpe ratios above use a constant 2% risk-free rate. The full-sample Sharpe ratios reported elsewhere use the actual time-varying T-bill rate (^IRX). The difference is material: the 2020–2021 split ran at near-zero rates while the 2023–2024 split ran at above 5%, so the two sets of numbers are not directly comparable on a like-for-like basis.

How Much Is the Model, and How Much Is the Optimiser?

One result in the walk-forward table deserves closer attention before we move on. The 3-Regime HMM with fixed weights achieved an average Sharpe of 1.06 — already a meaningful improvement over the 60/40 benchmark's 0.88. But the optimised variant pushed that to 1.27. The gap between those two numbers (0.21 Sharpe points) is attributable not to better regime detection, but to the mean-variance optimization layer sitting on top of the regime signal.

This decomposition is worth understanding before attributing all the outperformance to the Hidden Markov Model:

- **The regime model** tells you when to rotate — it identifies the market environment and adjusts the portfolio's risk posture accordingly.

- **The MVO layer** tells you how much to hold of each asset within a regime — it finds the efficient frontier given the regime-specific return and covariance estimates.

Both are doing real work. If you were implementing this in practice, I would suggest validating them separately: run the fixed-weight HMM first to confirm the regime signal is genuinely adding value, then layer in the optimisation to measure its incremental contribution. The regime detection gets you from 0.88 to 1.06; the optimization gets you from 1.06 to 1.27.

Honest Limitations

I want to be clear about what this analysis does and does not prove.

The period was favourable

2015–2024 rewarded the specific asset mix in this strategy. US equities had a dominant bull run, gold performed strongly in crisis periods, and Canadian bonds provided reasonable diversification until 2022. A strategy that holds meaningful gold allocation was structurally advantaged in this environment. A different 10-year window — say 2000–2010, which included two major bear markets and a decade of gold outperformance followed by a sharp reversal — would tell a different story.

The strategy's structural gold allocation deserves particular scrutiny. The 60/40 benchmark holds **zero gold**; this strategy holds 10–25% at all times. Over the 2015–2024 period, gold returned roughly 100%+, meaning some portion of the outperformance reflects a permanent tilt rather than regime-switching skill. A useful sanity check is to compare against a static 60/30/10 portfolio (equities/bonds/gold) that never switches regimes — if the gap narrows substantially, the active regime detection is doing less work than the headline numbers suggest. I have not run that comparison here, but it is the first thing a skeptical reader should do before drawing conclusions.

The crisis regime is thin

Only 52 of the 2,448 trading days across the full dataset were classified as crisis regime. This is too small a sample to draw strong conclusions about crisis-period behaviour, and it is why I hard-coded the crisis allocation rather than optimising it. If the next crisis looks structurally different from COVID which was a prolonged inflationary recession, for example, where bonds and equities fall together for an extended period the defensive allocation may not provide the protection expected.

Four walk-forward splits is the minimum viable test

Four one-year test periods gives us statistical coverage of meaningfully different market environments, but it is not a large sample. Each period is dominated by a single macro narrative, making it difficult to disentangle genuine strategy alpha from period-specific luck. The 3-Regime HMM Optimized outperformed in three of the four splits; the one period where it did not (Split 2, 2021–2022) is a reminder that no strategy is immune to a regime where the model's historical calibration breaks down.

Transaction costs are simplified

The backtest uses a fixed bid-ask spread cost of 0.08% per trade. In practice, ETF trading costs for retail investors in Canada vary by broker, lot size, and market conditions. The model counts 58 rebalances over 10 years, roughly one every two months, so costs are not a dominant factor (total simulated costs: $1,933 on $100,000 initial capital), but the simplification is worth noting.

Would I Actually Follow This Strategy?

Yes, with appropriate humility about the limitations described above. The core logic is sound: markets do exhibit persistent regimes, regime-aware allocation does reduce exposure to drawdowns, and the out-of-sample results hold up across meaningfully different market environments.

What I would not do is blindly follow the optimised weights output without understanding what is driving them. The strategy is not a black box — every weight decision can be traced back to a regime classification and a set of shrinkage-adjusted return estimates. When the model puts 85% in equities during a Bull regime, that is a coherent decision given low volatility, positive momentum, and tight credit spreads. When it rotates defensively, there is a reason.

The most important discipline for actually following this strategy is **not overriding the model when it makes you uncomfortable**. Regime-switching strategies derive their value from systematic rule-following. The moment you start second-guessing the signal because equities have been going up for six months and it feels wrong to hold 25% gold, you have undermined the entire premise.

The code, methodology, and full results are available on GitHub. If you find an error in the analysis or a better approach to any of the modelling decisions, I would genuinely like to hear it.