APPLIED MONTE CARLO 3

Monte Carlo Simulations for Agricultural Commodity Traders

Yield Risk, Seasonal Curves, Basis, and Crush Spreads in R

Ioanna Stamatiou

ODIN PRESS

Monte Carlo Simulations for Agricultural Commodity Traders

Paperback version, First edition: 2026.
ISBN: forthcoming (Bowker allocation pending)
Trim Size: 6.14 × 9.21 inch.

Source-code copyright. All R source code listings reproduced in this volume—including those in the Appendix, the chapter bodies, and any auxiliary file distributed alongside this book—are the copyrighted intellectual property of **Ioanna Stamatiou**. No reproduction, redistribution, adaptation, translation into another programming language, or incorporation into any derivative work, whether academic, commercial, or otherwise, is permitted without the prior written permission of the copyright holder. Single-machine execution by the individual purchaser of this volume for the sole purpose of reproducing a figure or table from this book is granted as a limited, non-transferable, non-sublicensable use and does not constitute redistribution.

Published by Odin Press.
`https://odinpress.org`
This book was typeset in LaTeX by Odin Press from a manuscript submitted by the author.

Printed and Bound in the United States of America.

9 8 7 6 5 4 3 2 1 0

Preface

Agricultural commodity traders run the oldest risk business in the world on some of the newest borrowed mathematics. The quantitative desks price corn with machinery built for equities and crude — flat-volatility diffusions, constant correlations, a single fair value quoted for new-crop six months before pollination — and the machinery is wrong in ways that are structured, seasonal, and expensive. A grain price has a calendar in its mean and a calendar in its volatility; it mean-reverts on the clock of the stocks-to-use cycle, not the trading day; it jumps when a heat ridge parks over the corn belt in July; and it cycles on its own supply response, because the acreage that answers this year's price is harvested into next year's market. None of that is exotic. All of it is absent from the standard toolkit, and every omission lands in the tail of somebody's book.

The literature has not closed the gap. Glasserman's *Monte Carlo Methods in Financial Engineering* is the method canon, calibrated to derivatives desks; Owen is comprehensive and audience-agnostic; the agricultural-economics literature understands cobwebs and storage but does not ship a simulation engine a desk can run. Schwartz–Smith and Sørensen supplied the factor models two decades ago, yet the trader who wants a working, calibrated, honest price simulator for corn — let alone for a book that spans corn, soybeans, wheat, and cocoa — still has to assemble it from a dozen papers and trust parameters nobody re-estimated on current data. This book is that assembly, done once, in the open, against data anyone can pull for free.

The book builds one engine in seven layers, every layer calibrated to real FRED and NOAA series with the estimating R code printed where it runs: a Fourier seasonal mean; a mean-reverting deviation with compound-Poisson weather jumps; a Schwartz–Smith two-factor equilibrium estimated by the Kalman filter; a seasonal stochastic-volatility envelope — thirteen percent annualized in February, thirty in July, the single most important ag-specific feature in the book — multiplied by a GARCH core; a Markov regime-switching layer with ENSO carried honestly as a weak, crop-specific drift covariate rather than the strong driver the data refuses to support; a competitive-storage cobweb that generates the multi-year cycles and asymmetric spikes from economics rather than statistics; and a demand layer whose elasticity, feed substitution, and ethanol-era energy coupling cut the supply-only one-percent tail from a fictitious ninety-one percent to the seventeen the market actually shows. The assembled engine then earns its keep: a storage option valued by least-squares Monte Carlo, the soybean board crush priced as the spread it is, a seasonal value-at-risk and expected shortfall computed month by month, and a backtest that reports its own failure — the Christoffersen independence rejection — as a finding instead of burying it. The last chapter recalibrates everything across four crops and reads off the parameter fingerprints that separate a storable temperate grain from a concentrated tropical perennial.

The chapters proceed in four movements. Chapter 1 states the case: a point forecast misstates downside risk, and only a distribution is honest. Chapter 2 builds the data anchor and the stylized-facts target list from the public series. Chapters 3 through 9 build the seven layers in order, each calibrated to the corn anchor and each leaving, explicitly, the residual the next layer models. Chapter 10 spends the engine on pricing and risk, and Chapter 11 is the comparative atlas — same structure, re-estimated parameters, from corn to cocoa. The appendix consolidates what a working desk actually photocopies: the full data ledger with retrieval code, the calibration atlas of every fitted parameter with its source section, and the Kupiec, Christoffersen, and Fisher test statistics with their decision rules. There is no code appendix because there is no code to hide: every script runs inside the chapter that uses it, against the cached data vintage that ships with the book.

I spent sixteen years in this market — five chartering Panamax grain cargoes out of Piraeus, eleven building the price simulation and the value-at-risk an oilseed desk in Geneva actually trusted — and the expensive lessons are all in here. The July that wasn't priced because the model treated volatility as constant through the pollination window. The crush hedge that bled through harvest because it was calibrated to the old-crop curve. The cocoa book that looked diversified until a single West African weather system proved it was one position in a trench coat. Each of those scars is now a layer, a test, or a fingerprint coordinate in this book. The standing instruction I give every new analyst is the book's whole thesis: price the distribution, not the number, and size for the tail you did not forecast.

Ioanna Stamatiou. Geneva, 2026.

Contents

Chapter 1

The Weather Lottery: Why Monte Carlo Is Mandatory in Agricultural Markets

The price of new-crop corn six months before pollination, P_{t+6}, is not a number anyone can know; it is a random variable whose distribution is set by a harvest nobody has seen yet. A broker who quotes a single fair value for it is quoting the mean of that distribution and quietly discarding everything that pays or destroys a position — the left tail, the seasonal vol ramp into July, the chance that one weather system over the corn belt moves the whole curve. The desks that survive size for the part they did not forecast, and to size for it they need the distribution, not the number.

This chapter makes the case that simulation is not a refinement in agricultural markets but the only honest instrument, and it delivers the case as runnable code. We state the agricultural price problem precisely; show why the closed-form pricing and value-at-risk formulas that work on equities quietly fail on crops; build the simplest Monte Carlo estimator and read its error bar; lay out the six-layer roadmap the rest of the book constructs; and pin every calibration in the book to free monthly FRED series the reader can pull in three lines. By the last section the reader has a corn-price simulation that prices a distribution instead of a number, and can see exactly how far a point forecast sits from the tail it ignores.

1.1 The agricultural price problem

Let P_t be the spot price of corn at month t. The structural fact that organizes this entire book is that P_t decomposes into a deterministic, calendar-locked component and a stochastic state that no forecast resolves until the harvest is in.

Write the log-price as the sum of a seasonal mean function and a stochastic deviation,

$$\log P_t = f(t) + X_t, \tag{1.1}$$

where $f(t)$ is periodic with a one-year period — the crop calendar, planting through harvest — and X_t is the unobserved state that carries the weather, the stocks-to-use buffer, and the supply response. This deterministic-plus-stochastic split is the standard starting point for commodity price modeling [1]. The seasonal mean $f(t)$ is knowable in advance: it repeats every crop year. The deviation X_t is not. A point forecast reports $\mathbb{E}[\log P_{t+h} \mid \mathcal{F}_t]$, the conditional mean of the sum, and in doing so it integrates X_{t+h} away. Everything a position can lose lives in the dispersion of X_{t+h}, which the conditional mean discards by construction. This is why a forecast and a risk number are different objects: the first is a point, the second is the spread of the random variable around it.

Conditioning on the information $\mathcal{F}_t$ available today makes the separation exact. Because $f(t+h)$ is deterministic and known, the only randomness left in $\log P_{t+h}$ is the state, so the forecast-error variance is entirely the conditional variance of X_{t+h},

$$\operatorname{Var}[\log P_{t+h} \mid \mathcal{F}_t] = \operatorname{Var}[X_{t+h} \mid \mathcal{F}_t]. \tag{1.2}$$

The right-hand side is the width of the band a position is actually exposed to, and the conditional mean a point forecast reports contains none of it. The whole task of the book is to model X_t richly enough that this conditional variance is estimated honestly, season by season, rather than collapsed to a single annual number.

The state X_t is not white noise; it has memory, and the dominant source of that memory is the stocks-to-use ratio — carryover inventory divided by annual consumption. A short crop draws stocks down, tightens the balance, and holds prices elevated until a later harvest rebuilds the buffer; an abundant crop does the reverse. That inventory buffer is why X_t mean-reverts toward the seasonal path rather than snapping back to it, and why the timing of the return is itself uncertain — a thin-stocks year can stay dislocated for many months. The persistence is a fundamental of the physical market, not a statistical artifact; Chapter 4 makes X_t an explicit mean-reverting process with weather-driven jumps.

The deterministic half is cheaper than the stochastic half, and its form is worth naming now even though Chapter 3 does the fitting. Because the crop calendar repeats every year, $f(t)$ is well represented by a low-order Fourier series in the month-of-year,

$$f(t) = a_0 + \sum_{k=1}^{K} \big(a_k \cos(2\pi k\, \tau_t) + b_k \sin(2\pi k\, \tau_t)\big), \tag{1.3}$$

where $\tau_t \in [0, 1)$ is the fractional position within the crop year and K is small — two or three harmonics already capture the planting-trough-to-harvest-peak shape. A handful of coefficients, fit once, are reused on every simulated path; all the modeling effort then goes where the risk is, into the law of X_t. The two-component split here is the seed of the layered model of §1.4, and Chapter 2 turns $f(t)$ and the law of X_t into quantities estimated from data.

The thirty-four-year record makes the point concretely. Figure 1.1 plots the FRED global corn price [2] monthly from 1992 to 2026, from a low near 75 USD per metric ton in the late-1990s surplus years to a peak of 348 USD per metric ton in 2022. Three episodes are shaded. The 2008 food-price crisis carried corn to 287 USD per metric ton; the 2010–2012 run, capped by the 2012 US Midwest drought, reached 333 USD per metric ton in the pollination-window heat of that summer; and the 2022 Ukraine grain shock took the series to its record. None of the three was in any point forecast made the prior winter. Each was a draw from the tail of X_t — a weather system or a war, not a revision to the seasonal mean. A trader long new-crop corn into any of those summers discovered the distribution the expensive way.

The 2012 episode is worth one more look because it is the cleanest weather draw in the record. Corn entered the year near 260 USD per metric ton on a normal balance sheet; the winter point forecasts had the new crop drifting sideways into autumn on the usual seasonal mean. Then the rain stopped. A rain-deficit ridge parked over the corn belt through July's pollination window, the USDA cut its yield estimate sharply in successive WASDE reports, and the series ran to 333 USD per metric ton — a 28% move in a single growing season, with essentially none of it in the deterministic $f(t)$. A position sized off the winter forecast's narrow band was carried straight through it. This is the mechanism the whole book exists to price: the loss came not from the mean being wrong but from the dispersion of X_t being unmodeled.

Chapter 3

The Crop Calendar: Deterministic Seasonality in the Price Mean (Layer 1)

Chapter 2 measured the stylized facts and put a seasonal signature at the head of the list. This chapter builds the first and most deterministic layer of the simulation engine: the seasonal mean, the calendar-locked component of the log-price $\log P_t$ that repeats every crop year while the stochastic deviation X_t does not. It is the function $f(t)$ in Chapter 1's decomposition $\log P_t = f(t) + X_t$, the average shape agricultural prices carry across the year. The crop calendar of planting, growing, and harvest imposes a shape on the average price across the year that repeats, that is exploitable, and that a flat model throws away.

The layer is cheap and the payoff is structural. A low-order Fourier series in the month-of-year, the function $f(t)$ of Chapter 1, captures the annual cycle in a handful of coefficients, and fitting it by least squares to the corn and wheat series yields an interpretable amplitude and phase — how big the seasonal swing is and when it peaks. We motivate the calendar as a genuine signal rather than noise; write the trigonometric seasonal mean and choose its order; fit it to corn and wheat; show how the mean splices two crop years into the harvest break that flat models miss; and read the residual diagnostics that hand the leftover stochastic deviation to the next layer. The deliverable is a calibrated Fourier seasonal-mean model for corn and wheat, and the residual it leaves behind is the input to Chapter 4.

3.1 The crop calendar as signal

The seasonal mean is the function $f(t)$ in the decomposition $\log P_t = f(t) + X_t$ of Chapter 1, and the claim of this section is that $f(t)$ is a real, estimable signal rather than an artifact of a noisy series.

Set up the estimand precisely. The seasonal mean is the expected log-price as a function of the month-of-year, after removing the slow trend,

$$f(t) = \mathbb{E}[\log P_t \mid m(t)] - \mathrm{trend}(t), \tag{3.1}$$

with $m(t) \in \{1, \dots, 12\}$ the calendar month. It is deterministic because the crop calendar is deterministic: planting, pollination, and harvest happen at roughly the same dates every year, and they move physical supply onto and off the market on a fixed annual schedule. That schedule shows up as a repeating shape in the average price — high when old-crop stocks are tight before a harvest, low when the new crop floods in — and because it repeats, it is the one component of the price a model can know in advance [8]. The claim that $f(t)$ is estimable rests on repetition. Any single crop year offers only twelve monthly prices, hopelessly confounded with that year's idiosyncratic weather and macro shocks; but stacking thirty-four crop years and asking what recurs at each calendar position averages those idiosyncratic shocks toward zero and leaves the systematic calendar shape standing. This is identification by replication, and it is why a decades-long series is a requirement, not a luxury: the seasonal amplitude is a few percent while the year-to-year noise is twenty, so it takes many stacked years before the signal emerges at the precision a calibration needs. The trend subtraction in the definition matters for the same reason — without it, the slow climb of the nominal price across the decades would leak into the seasonal estimate, since months falling later in the sample sit at a higher level; removing a smooth trend first measures the month-of-year effect as a deviation from the local level rather than from where in the thirty-four years the month happened to land.

```
suppressMessages(library(ggplot2))
set.seed(20260611)

FILES <- c(corn = "PMAIZMTUSDM_corn.csv", soy = "
    PSOYBUSDM_soybeans.csv",
           wheat = "PWHEAMTUSDM_wheat.csv", cocoa = "
```

```
               PCOCOUSDM_cocoa.csv",
           coffee = "PCOFFOTMUSDM_coffee_arabica.csv",
               sugar = "PSUGAISAUSDM_sugar.csv")
load_crop <- function(name) {
  d <- read.csv(file.path("book_code", "data", FILES[[name
      ]]), stringsAsFactors = FALSE)
  names(d) <- c("date", "value"); d$date <- as.Date(d$date)
  d <- d[is.finite(d$value), ]; d[order(d$date), ]
}

# Seasonal-mean fit: log-price on a quadratic trend plus
    two annual harmonics.
seas_fit <- function(d) {
  t <- seq_len(nrow(d)); m <- as.integer(format(d$date, "%m
      "))
  lm(log(d$value) ~ t + I(t^2) + sin(2*pi*m/12) + cos(2*pi*
      m/12)
                              + sin(4*pi*m/12) + cos(4*pi*m
                                  /12))
}
```

Figure 3.1 overlays the fitted seasonal mean on the realized corn price across the full window. The fitted curve — quadratic trend plus two harmonics — tracks the broad sweep of the series and carries a visible annual ripple, while the realized price oscillates around it with the large stochastic excursions of the weather years. The fit explains about 51% of the variance of the log-price for corn; the bulk of that is the trend, and the seasonal ripple is a few percent on top, but it is a *systematic* few percent that recurs every year rather than averaging out.

```
# Figure 3.1: fitted seasonal mean over the realized corn
    series.
d <- load_crop("corn"); d$fitted <- exp(predict(seas_fit(d)
    ))
ld <- rbind(data.frame(date = d$date, value = d$value,
    series = "realized"),
            data.frame(date = d$date, value = d$fitted,
                series = "seasonal mean"))
ggplot(ld, aes(date, value, linetype = series)) + geom_line
    (linewidth = 0.55) +
  labs(x = NULL, y = "corn price (USD/t)") + theme_bw()
```

A single fair value for new-crop corn, quoted six months before pollination, is a guess wearing a tie. Agricultural prices carry a calendar in their mean, a calendar in their volatility, weather jumps, climate regimes, and a supply cycle of their own making — and the standard quantitative toolkit, built for equities and crude, misses every one of them.

This book builds the missing engine in seven calibrated layers: a seasonal mean, mean-reverting weather jumps, a Schwartz–Smith equilibrium, the seasonal volatility envelope that runs from 13% in February to 30% in July, climate regimes, the competitive-storage cobweb, and the demand layer that cuts a supply-only tail by four-fifths. Every layer is fitted to free public FRED and NOAA data with the R code printed where it runs, then spent on what a desk actually needs: storage options by least-squares Monte Carlo, the soybean crush spread, month-by-month value-at-risk and expected shortfall, and an honestly reported backtest. A closing atlas recalibrates the whole engine from corn to cocoa and reads each crop's character straight off its parameter fingerprint.

Written by a quant who learned distributions the expensive way, for the trader who has to run the model. Price the distribution, not the number — and size for the tail you did not forecast.