Problems in High-Frequency Econometrics

Market Microstructure, Adverse Selection, and Price Impact

Aryan Ayyar

Version: June 2026

Working draft for study, criticism, and correction

I have tried my best to make the arguments, references, and calculations as accurate as possible, but any manuscript that claims perfection should probably be shorted. Corrections are very welcome. So is criticism, especially the useful kind: places where an explanation is unclear, a problem is too abrupt, a reference could be better, a derivation has gone suspiciously fast, or a joke has committed a small academic crime. Readers are warmly invited to write to me with what they liked, what they disliked, and what should be fixed in the next version.

Preface

These notes began as a private attempt to make market microstructure stop behaving like a fog machine. Every model looked innocent at first: a spread here, a trade sign there, one little conditional expectation minding its own business. Then, three lines later, the notation had formed a small government and was demanding tribute. So the book is organized as solved problems. Not because solved problems are the only way to learn, but because they force the model to show its hands. A transaction-price equation has to reveal its autocovariances. A Bayesian market maker has to admit what is being inferred from order flow. A Kyle insider has to account for camouflage, price impact, timing, disclosure, liquidity, and the mildly inconvenient fact that other people are also thinking. The goal is to bridge the old machinery with the new questions. Roll, MRR, Glosten-Milgrom, sequential trade, and Kyle-style strategic trading sit next to problems on adverse selection, order-flow persistence, information sharing, risk aversion, stochastic liquidity, machine-learned equilibria, and market impact. The attitude I want for the reader is *shoshin*: beginner's mind. Not the fake beginner's mind where one pretends not to know things. The useful one: read the equation as if it can still surprise you. Ask what is observed, what is latent, what is being projected, what is being conditioned on, and which symbol is quietly carrying the entire economic assumption. The craft standard is *takumi*: patient, exacting workmanship. Do not wave at a covariance. Compute it. Do not call something "liquidity" until you know whether it means spread, depth, resilience, noise-trader variance, or a PowerPoint slide having a difficult week. Do not worship elegance either. Some markets are elegant; many are just noisy auctions wearing a tie. This is not meant to be a sacred text. Please write in the margins. Argue with the problems. Circle suspicious assumptions. Put a small question mark next to any model that seems too pleased with itself. I like this field because it sits at a strange intersection: probability, econometrics, game theory, institutional detail, and the ancient human desire to buy before other people and sell before they notice. It is mathematical, but not sterile. It is empirical, but not obedient. The data are high frequency; the mistakes are also high frequency. The three chapters move in a deliberate arc. Chapter 1 starts with prices and trade signs: bid-ask bounce, permanent impact, delayed adjustment, and latent efficient prices. Chapter 2 moves to adverse selection: Bayesian updating, sequential trade, PIN, VPIN, and modern diagnostics. Chapter 3 then turns to Kyle-style equilibrium and market impact: strategic informed trading, dynamic revelation, disclosure, stochastic liquidity, neural learning, execution, and metaorders. The prerequisites are not mysterious, but they are real. You should be comfortable with probability, conditional expectation, covariance algebra, linear regression, likelihoods, basic time series, and the idea that filtrations are just information sets wearing formal clothing. A first course in asset pricing or econometrics is enough for the core problems; stochastic calculus, point processes, optimization, and game theory become useful as the notes move into dynamic Kyle, sequential trade, and execution. The most important prerequisite, though, is tolerance for being confused in a productive way. Use the problem labels as a map, not a caste system. Problems marked **Core** should be solvable after a first graduate course in asset pricing or econometrics. Problems marked **Advanced** ask for stronger time-series, stochastic-process, optimization, or game-theory tools. Problems marked **Research** are meant for reading groups, replication exercises, thesis exploration, and the occasional afternoon in which someone says, with catastrophic confidence, "this should be quick." Each problem is designed to teach either a model, an identification argument, or an econometric habit;

the solutions are not there so the reader can admire them from a respectful distance, but so the reader can reverse-engineer the moving parts and then stress-test them. A small warning: if you try to memorize these models, they will become very polite and completely useless. Work through them instead. Change an assumption. Break a coefficient. Ask what the market maker sees. Ask what the insider hides. Ask what the econometrician can identify after everyone has finished being strategic. If these notes work, they should leave the reader with two instincts: first, that every microstructure model is a disciplined story about who knows what and when; second, that the algebra is not decoration. It is where the story either survives or quietly falls down the stairs. That is the spirit of the notes: beginner's mind, craftsman's patience, and just enough suspicion to keep the equations honest.

Notation and Difficulty Guide

The notation is mostly local to each problem, but the following symbols recur often enough that it is worth fixing the reader's eye before the algebra starts. When a chapter uses a symbol differently, the local definition in that problem or section wins.

Symbol	Common meaning in these notes
t, τ, T	calendar time, trade time, or terminal date, depending on the model context
p_t, P_t	observed transaction price or quoted market price
m_t	latent efficient price, conditional fundamental value, or martingale price component
r_t, Δp_t	return or transaction-price change
q_t, ϵ_t	signed trade indicator, usually $+1$ for buyer-initiated and -1 for seller-initiated trades
s, c	spread, half-spread, or temporary bid–ask component; always check the local normalization
λ	price-impact coefficient, Kyle lambda, or slope of price on order flow
x, X, θ	informed-trader order, strategy, or trading rate
u, y, Y	liquidity/noise-trader order or cumulative noise order flow
v, z	asset value, terminal payoff, or fundamental signal
Q, Y, $X+Y$	aggregate order flow observed by the market maker
μ, σ^2, Σ	mean, variance, or posterior variance
$\mathcal{F}_t$	information available up to time t
N_t, λ_t	counting process and arrival intensity in sequential-trade or point-process models
α, β, γ	generic model coefficients; their meaning is deliberately local
$\mathbb{E}[\cdot]$, $\mathrm{Var}(\cdot)$, $\mathrm{Cov}(\cdot,\cdot)$	expectation, variance, and covariance under the probability model in force

Core problems are the main road. They usually require probability, conditional expectation, covariance algebra, regression logic, and patience with notation. A reader should be able to solve them after a first serious course in asset pricing, econometrics, or probability.

Advanced problems ask for more technique: likelihood manipulation, state-space or regime models, stochastic-process intuition, execution optimization, or equilibrium reasoning. These problems are not optional decoration; they are where the model starts showing its machinery.

Research problems are exploratory. They may ask for simulation design, model comparison, diagnostics, robustness, or extensions from recent papers. They are meant to be read with a pencil, a browser tab, and some emotional preparation. The aim is not to finish them quickly; the aim is to learn what would have to be true for the argument to survive contact with data or with a slightly more annoying model.

Contents

Chapter 1

Foundations: Roll, MRR, and Price Discovery

This opening part builds the algebraic core of high-frequency econometrics. The reader should become comfortable moving between a transaction-price model, its return autocovariances, its economic interpretation, and the estimator implied by those moment restrictions.

Notation for Part I

Symbol	Meaning
m_t	latent efficient price or fundamental value
p_t	observed transaction price
$q_t \in \{-1, +1\}$	trade direction, with +1 for buyer-initiated trades
c	half-spread or transitory bid-ask component
u_t	efficient-price innovation
λ	permanent price impact of signed order flow
α	Markov continuation probability for trade signs
ϕ	trade-sign persistence coefficient, often $\phi = 2\alpha - 1$
Δp_t	transaction-price change, $p_t - p_{t-1}$

1.1 Trade Signs, Bid-Ask Bounce, and Serial Dependence

Problem 1.1 (Core: Markov trade signs and uncorrelated innovations)**.** Let the trade direction process satisfy $q_t \in \{-1, +1\}$ and

$$\mathbb{P}(q_{t+1} = +1 \mid q_t = +1) = \mathbb{P}(q_{t+1} = -1 \mid q_t = -1) = \alpha,$$

with

$$\mathbb{P}(q_{t+1} = +1 \mid q_t = -1) = \mathbb{P}(q_{t+1} = -1 \mid q_t = +1) = 1 - \alpha.$$

Assume the stationary distribution is symmetric, so $\mathbb{P}(q_t = 1) = \mathbb{P}(q_t = -1) = 1/2$.

(a) Compute the eight path probabilities for (q_t, q_{t+1}, q_{t+2}).

(b) Show that the linear projection coefficient of q_{t+1} on q_t is

$$\phi = \frac{\mathrm{Cov}(q_{t+1}, q_t)}{\mathrm{Var}(q_t)} = 2\alpha - 1.$$

(c) Define the innovation $v_{t+1} = q_{t+1} - \phi q_t$. Verify that $\mathbb{E}[v_{t+1}] = 0$ and $\mathrm{Cov}(v_{t+1}, v_t) = 0$.

(d) Explain why the innovations are uncorrelated but not independent.

Solution.

The path probabilities follow by multiplying the stationary initial probability by the two transition probabilities:

Path	q_t	q_{t+1}	q_{t+2}	Probability
1	+1	+1	+1	$\frac{1}{2}\alpha^2$
2	+1	+1	−1	$\frac{1}{2}\alpha(1-\alpha)$
3	+1	−1	+1	$\frac{1}{2}(1-\alpha)^2$
4	+1	−1	−1	$\frac{1}{2}\alpha(1-\alpha)$
5	−1	−1	−1	$\frac{1}{2}\alpha^2$
6	−1	−1	+1	$\frac{1}{2}\alpha(1-\alpha)$
7	−1	+1	−1	$\frac{1}{2}(1-\alpha)^2$
8	−1	+1	+1	$\frac{1}{2}\alpha(1-\alpha)$

Because $q_t^2 = 1$, $\mathrm{Var}(q_t) = 1$. Also,

$$\mathbb{E}[q_{t+1}q_t] = \mathbb{P}(q_{t+1} = q_t) - \mathbb{P}(q_{t+1} \neq q_t) = \alpha - (1-\alpha) = 2\alpha - 1.$$

Thus $\phi = 2\alpha - 1$.

Now

$$\mathbb{E}[v_{t+1}] = \mathbb{E}[q_{t+1} - \phi q_t] = 0.$$

Similarly,

$$\mathrm{Cov}(v_{t+1}, v_t) = \mathbb{E}[(q_{t+1} - \phi q_t)(q_t - \phi q_{t-1})].$$

Using stationarity and the Markov transition structure,

$$\mathbb{E}[q_{t+1}q_t] = \phi, \qquad \mathbb{E}[q_t q_{t-1}] = \phi, \qquad \mathbb{E}[q_{t+1}q_{t-1}] = \phi^2,$$

so

$$\mathrm{Cov}(v_{t+1}, v_t) = \phi - \phi\phi^2 - \phi + \phi^3 = 0.$$

The innovations are not independent because the process is binary and Markov. Once q_t is known, $v_{t+1} = q_{t+1} - \phi q_t$ can only take two values whose probabilities depend on the current state. A useful exercise is to compute a higher-order moment such as $\mathbb{E}[v_{t+1}^2 v_t]$ or $\mathrm{Cov}(v_{t+1}^2, v_t^2)$ and verify that it does not factor as it would under independence. □

Takeaway. *Trade-sign innovations can be martingale differences without being independent. This distinction matters because many high-frequency estimators use second moments, while likelihood and filtering arguments need the full dependence structure.*

References for this problem.

- Madhavan, Richardson, and Roomans (1997), for signed order flow and transaction-level price changes.
- Hasbrouck (2007), for the empirical microstructure treatment of trade indicators and transaction prices.
- Lee and Ready (1991), for trade-direction classification in intraday data.

1.2 Roll Spread Identification

Problem 1.2 (Core: Roll's covariance restriction)**.** Consider the Roll transaction-price model

$$p_t = m_t + cq_t, \qquad m_t = m_{t-1} + u_t,$$

where $q_t \in \{-1, +1\}$ is iid with $\mathbb{E}[q_t] = 0$ and $\mathrm{Var}(q_t) = 1$, and u_t is iid with mean zero, variance σ_u^2, and independent of all trade signs.

(a) Derive Δp_t.

(b) Compute $\mathrm{Var}(\Delta p_t)$.

(c) Compute $\mathrm{Cov}(\Delta p_t, \Delta p_{t-1})$ and $\mathrm{Cov}(\Delta p_t, \Delta p_{t-k})$ for $k > 1$.

(d) Show how the half-spread c is identified from the first return autocovariance.

(e) Explain why transaction returns are negatively autocorrelated even when efficient-price returns are serially uncorrelated.

Solution.

The return is

$$\Delta p_t = p_t - p_{t-1} = u_t + c(q_t - q_{t-1}).$$

Since u_t is independent of trade signs,

$$\operatorname{Var}(\Delta p_t) = \sigma_u^2 + c^2 \operatorname{Var}(q_t - q_{t-1}).$$

With iid signs and unit variance,

$$\operatorname{Var}(q_t - q_{t-1}) = \operatorname{Var}(q_t) + \operatorname{Var}(q_{t-1}) = 2,$$

so

$$\operatorname{Var}(\Delta p_t) = \sigma_u^2 + 2c^2.$$

For the first autocovariance,

$$\operatorname{Cov}(\Delta p_t, \Delta p_{t-1}) = \operatorname{Cov}(u_t + c(q_t - q_{t-1}), u_{t-1} + c(q_{t-1} - q_{t-2})).$$

All terms involving u_t, u_{t-1} and non-overlapping signs vanish, leaving

$$c^2 \operatorname{Cov}(q_t - q_{t-1}, q_{t-1} - q_{t-2}) = c^2(-\operatorname{Var}(q_{t-1})) = -c^2.$$

For $k > 1$, there is no shared trade-sign term and no serial correlation in u_t, so

$$\operatorname{Cov}(\Delta p_t, \Delta p_{t-k}) = 0.$$

Thus

$$c = \sqrt{-\operatorname{Cov}(\Delta p_t, \Delta p_{t-1})},$$

provided the first autocovariance is negative.

The negative autocovariance is bid-ask bounce. A buyer-initiated trade moves the transaction price up by c relative to the efficient price. If the next trade is seller-initiated, the transaction price mechanically moves back down by $2c$ even when the efficient price has no predictable component. □

Takeaway. *Roll identification is a moment argument: the same transitory spread component appears with opposite signs in two adjacent returns. This creates negative first-order return autocovariance and identifies the half-spread in the ideal iid trade-sign model.*

References for this problem.

- Roll (1984), for the covariance restriction and implicit effective spread estimator.
- Hasbrouck (2007), for a clean textbook exposition of bid-ask bounce.
- Harris (1990), for finite-sample behavior of the Roll estimator.

Problem 1.3 (Core: Persistent trade signs and biased Roll estimates)**.** Keep the Roll price equation $p_t = m_t + cq_t$ and $m_t = m_{t-1} + u_t$, but now allow trade signs to be weakly persistent:

$$\mathbb{E}[q_t] = 0, \qquad \mathrm{Var}(q_t) = 1, \qquad \mathrm{Cov}(q_t, q_{t-1}) = \rho,$$

with $\mathrm{Cov}(q_t, q_{t-k}) = 0$ for $k > 1$. Assume u_t remains independent white noise.

(a) Derive $\mathrm{Var}(\Delta p_t)$.

(b) Derive $\mathrm{Cov}(\Delta p_t, \Delta p_{t-1})$ and $\mathrm{Cov}(\Delta p_t, \Delta p_{t-2})$.

(c) If a researcher wrongly applies the iid Roll estimator

$$\widehat{c}_{\text{Roll}} = \sqrt{-\widehat{\gamma}_1}, \qquad \gamma_1 = \mathrm{Cov}(\Delta p_t, \Delta p_{t-1}),$$

when is the estimator real?

(d) Interpret the economic bias.

Solution.

As before,

$$\Delta p_t = u_t + c(q_t - q_{t-1}).$$

The variance is

$$\mathrm{Var}(\Delta p_t) = \sigma_u^2 + c^2\,\mathrm{Var}(q_t - q_{t-1}) = \sigma_u^2 + c^2(1 + 1 - 2\rho),$$

so

$$\mathrm{Var}(\Delta p_t) = \sigma_u^2 + 2c^2(1 - \rho).$$

For the first autocovariance,

$$\begin{aligned}
\gamma_1 &= c^2\,\mathrm{Cov}(q_t - q_{t-1}, q_{t-1} - q_{t-2}) \\
&= c^2\{\mathrm{Cov}(q_t, q_{t-1}) - \mathrm{Cov}(q_t, q_{t-2}) - \mathrm{Var}(q_{t-1}) + \mathrm{Cov}(q_{t-1}, q_{t-2})\} \\
&= c^2(\rho - 0 - 1 + \rho) = -c^2(1 - 2\rho).
\end{aligned}$$

For the second autocovariance,

$$\begin{aligned}\gamma_2 &= c^2 \operatorname{Cov}(q_t - q_{t-1}, q_{t-2} - q_{t-3}) \\ &= c^2\{0 - 0 - \rho + 0\} = -c^2\rho.\end{aligned}$$

Higher autocovariances are zero under the assumed one-lag sign dependence.

The naive Roll estimator uses

$$\widehat{c}^2_{\text{Roll}} = -\gamma_1 = c^2(1 - 2\rho).$$

It is real only when $\rho \leq 1/2$. For $0 < \rho < 1/2$ it is biased downward:

$$c_{\text{Roll}} = c\sqrt{1 - 2\rho} < c.$$

For $\rho > 1/2$, the first return autocovariance becomes positive and the classic Roll square-root formula breaks down. □

Takeaway. *Positive order-flow persistence weakens or overturns bid-ask bounce in returns. The Roll estimator is not merely biased under persistent trade signs; it can become undefined because the identifying sign of the first autocovariance is lost.*

References for this problem.

- Roll (1984), for the iid-trade-sign benchmark.
- Madhavan, Richardson, and Roomans (1997), for persistent signed order flow and permanent price impact.
- Lee and Ready (1991), for empirical trade-sign construction.

1.3 Permanent Impact and Informational Order Flow

Problem 1.4 (Core: Correlated fundamentals and upward spread bias)**.** Suppose transaction prices still satisfy

$$p_t = m_t + cq_t,$$

but the efficient-price innovation is correlated with contemporaneous trade direction:

$$m_t = m_{t-1} + u_t, \qquad \mathbb{E}[u_t] = 0, \qquad \operatorname{Var}(u_t) = \sigma_u^2,$$

$$\operatorname{Cov}(u_t, q_t) = \rho\sigma_u, \qquad \operatorname{Cov}(u_t, q_s) = 0 \quad \text{for } s \neq t,$$

where q_t is iid with mean zero and unit variance.

(a) Compute $\mathrm{Var}(\Delta p_t)$.

(b) Compute $\mathrm{Cov}(\Delta p_t, \Delta p_{t-1})$.

(c) Show how the naive Roll estimator is biased.

(d) Interpret the sign of the bias in terms of adverse selection.

Solution.

We again have

$$\Delta p_t = u_t + c(q_t - q_{t-1}).$$

The variance is

$$\begin{aligned}\mathrm{Var}(\Delta p_t) &= \mathrm{Var}(u_t) + c^2\,\mathrm{Var}(q_t - q_{t-1}) + 2c\,\mathrm{Cov}(u_t, q_t - q_{t-1}) \\ &= \sigma_u^2 + 2c^2 + 2c\rho\sigma_u.\end{aligned}$$

The first autocovariance is

$$\begin{aligned}\gamma_1 &= \mathrm{Cov}(u_t + c(q_t - q_{t-1}), u_{t-1} + c(q_{t-1} - q_{t-2})) \\ &= c\,\mathrm{Cov}(q_t - q_{t-1}, u_{t-1}) + c^2\,\mathrm{Cov}(q_t - q_{t-1}, q_{t-1} - q_{t-2}).\end{aligned}$$

The first term equals $-c\rho\sigma_u$ because q_{t-1} is correlated with u_{t-1}. The second term is $-c^2$. Hence

$$\gamma_1 = -c(c + \rho\sigma_u).$$

The naive Roll estimator gives

$$c_{\mathrm{Roll}} = \sqrt{-\gamma_1} = \sqrt{c(c + \rho\sigma_u)}.$$

If $\rho > 0$, then $c_{\mathrm{Roll}} > c$.

The reason is economic. A buy order is now associated not only with paying the ask side of the spread, but also with a positive efficient-price innovation. The transaction-price reversal pattern reflects both bid-ask bounce and adverse selection. A purely transitory Roll interpretation over-attributes this effect to the spread. □

Takeaway. *When order flow is informative, transaction-price autocovariances mix transitory spread effects with permanent information effects. A negative first autocovariance no longer identifies only the mechanical half-spread.*

References for this problem.

- Glosten and Milgrom (1985), for Bayesian quote setting under asymmetric information.

- Madhavan, Richardson, and Roomans (1997), for reduced-form permanent price impact.
- Glosten and Harris (1988), for decomposing bid-ask spread components.

Problem 1.5 (Advanced: MRR permanent and transitory components)**.** Consider the Madhavan-Richardson-Roomans style model

$$p_t = m_t + cq_t, \qquad m_t = m_{t-1} + \lambda q_t + u_t,$$

where q_t is iid with mean zero and unit variance, and u_t is iid with variance σ_u^2 and independent of trade signs.

(a) Derive Δp_t.

(b) Compute $\mathrm{Var}(\Delta p_t)$ and $\mathrm{Cov}(\Delta p_t, \Delta p_{t-1})$.

(c) Explain how the autocovariance differs from the pure Roll model.

(d) Suppose σ_u^2 is known. Can c and λ be separately identified from $\gamma_0 = \mathrm{Var}(\Delta p_t)$ and $\gamma_1 = \mathrm{Cov}(\Delta p_t, \Delta p_{t-1})$?

Solution.

Since

$$p_t = m_t + cq_t = m_{t-1} + \lambda q_t + u_t + cq_t,$$

we obtain

$$\Delta p_t = u_t + (\lambda + c)q_t - cq_{t-1}.$$

The variance is

$$\gamma_0 = \mathrm{Var}(\Delta p_t) = \sigma_u^2 + (\lambda + c)^2 + c^2.$$

The first autocovariance is

$$\begin{aligned}\gamma_1 &= \mathrm{Cov}(u_t + (\lambda + c)q_t - cq_{t-1}, u_{t-1} + (\lambda + c)q_{t-1} - cq_{t-2}) \\ &= -c(\lambda + c),\end{aligned}$$

and higher autocovariances are zero.

Relative to Roll, the negative first autocovariance is no longer $-c^2$. Instead, it is $-c(\lambda+c)$. A trade has a permanent component λq_t and a transitory component cq_t. The next return reverses only the transitory part, but the covariance loads on the total contemporaneous trade-related movement $\lambda + c$.

If σ_u^2 is known, define

$$A = \gamma_0 - \sigma_u^2 = (\lambda + c)^2 + c^2, \qquad B = -\gamma_1 = c(\lambda + c).$$

Let $x = \lambda + c$. Then

$$A = x^2 + c^2, \qquad B = cx.$$

Therefore

$$(x + c)^2 = A + 2B, \qquad (x - c)^2 = A - 2B.$$

With sign restrictions such as $c > 0$ and $x = \lambda + c > 0$, one can recover

$$x = \frac{\sqrt{A + 2B} + \sqrt{A - 2B}}{2}, \qquad c = \frac{\sqrt{A + 2B} - \sqrt{A - 2B}}{2},$$

and then $\lambda = x - c$. Without economically meaningful sign restrictions, the moment equations can have observationally equivalent sign choices. □

Takeaway. *MRR separates price movements into permanent information effects and transitory trading-cost effects. The same autocovariance that identifies the Roll spread now identifies a product of permanent and transitory components, so interpretation requires a richer model. This is also the first bridge to Chapter 2: the reduced-form object*

$$\gamma_1 = -c(\lambda + c)$$

summarizes an average permanent response to order flow. In a structural sequential-trade model, that permanent response is generated by Bayesian belief updating, informed-arrival intensities, and the current posterior state rather than by a constant primitive coefficient.

References for this problem.

- Madhavan, Richardson, and Roomans (1997), for the permanent/transitory transaction-price decomposition.
- Hasbrouck (2007), for price-impact regressions and trade-correlated price changes.
- Glosten and Milgrom (1985), for the structural Bayesian interpretation behind permanent price response.

1.4 Finite-Sample Bias and Structural Identification

Problem 1.6 (Advanced: Harris-Jensen bias in the Roll estimator)**.** Suppose the first-order transaction-return covariance is estimated from a finite sample and satisfies the approximation

$$\widehat{\gamma}_1 \approx N(-c^2, \sigma_\gamma^2).$$

The Roll half-spread estimator is

$$\widehat{c}_R = \sqrt{-\widehat{\gamma}_1}$$

when $\widehat{\gamma}_1 < 0$. Some empirical studies discard observations with $\widehat{\gamma}_1 \geq 0$; others set the spread estimate to zero.

(a) Use a second-order Taylor expansion to approximate $\mathbb{E}[\sqrt{-\widehat{\gamma}_1}]$ around $\gamma_1 = -c^2$.

(b) Explain the sign of the Jensen bias.

(c) Discuss why the truncation rule $\widehat{c}_R = 0$ when $\widehat{\gamma}_1 \geq 0$ creates a separate finite-sample issue.

(d) Why is this problem more severe for liquid stocks with small spreads?

Solution.

Let

$$g(x) = \sqrt{-x}, \qquad x < 0.$$

Then

$$g'(x) = -\frac{1}{2(-x)^{1/2}}, \qquad g''(x) = -\frac{1}{4(-x)^{3/2}}.$$

A second-order expansion around $\gamma_1 = -c^2$ gives

$$\mathbb{E}[g(\widehat{\gamma}_1)] \approx g(\gamma_1) + \frac{1}{2}g''(\gamma_1)\,\mathrm{Var}(\widehat{\gamma}_1).$$

Since $g(-c^2) = c$ and

$$g''(-c^2) = -\frac{1}{4c^3},$$

we obtain

$$\mathbb{E}[\widehat{c}_R] \approx c - \frac{\sigma_\gamma^2}{8c^3}.$$

The curvature term is negative, so the square-root transformation generates downward Jensen bias conditional on remaining in the negative-covariance region.

The truncation rule is different from this curvature bias. If $\widehat{\gamma}_1$ is positive, the classical Roll square root is not real. Setting the estimate to zero piles mass at zero; discarding the observation creates selection on the sign of the sample covariance. Both practices change the sampling distribution. The issue is especially severe when c is small relative to σ_γ, because then finite samples frequently put $\widehat{\gamma}_1$ on the wrong side of zero. □

Takeaway. *Roll's identity is exact in the ideal model, but the estimator can be fragile. The square-root map is nonlinear, and the empirical convention for nonnegative sample autocovariances is itself an econometric choice.*

References for this problem.

- Harris (1990), for statistical properties and finite-sample bias in the Roll estimator.

- Roll (1984), for the population covariance restriction.
- Hasbrouck (2007), for practical limitations of Roll-style estimation.

Problem 1.7 (Core: MRR identification from regression coefficients)**.** Consider the MRR model written as

$$p_t = \mu_t + \phi x_t, \qquad \mu_t = \mu_{t-1} + \theta(x_t - \rho x_{t-1}) + \epsilon_t,$$

where $x_t \in \{-1, +1\}$ is the trade sign and

$$\mathbb{E}[x_t \mid x_{t-1}] = \rho x_{t-1}.$$

(a) Derive a return equation for Δp_t as a function of x_t and x_{t-1}.

(b) Suppose the researcher estimates

$$\Delta p_t = a x_t + b x_{t-1} + u_t.$$

Assuming ρ is known or separately estimated, recover θ and ϕ.

(c) Explain why identification becomes weak as $\rho \to 1$.

(d) Interpret θ and ϕ economically.

Solution.

The transaction-price change is

$$\begin{aligned}\Delta p_t &= (\mu_t - \mu_{t-1}) + \phi x_t - \phi x_{t-1} \\ &= \theta(x_t - \rho x_{t-1}) + \epsilon_t + \phi x_t - \phi x_{t-1} \\ &= (\phi + \theta) x_t - (\phi + \rho\theta) x_{t-1} + \epsilon_t.\end{aligned}$$

Thus the population regression coefficients satisfy

$$a = \phi + \theta, \qquad b = -(\phi + \rho\theta).$$

Adding the two equations gives

$$a + b = (1 - \rho)\theta,$$

so

$$\theta = \frac{a + b}{1 - \rho}, \qquad \phi = a - \theta.$$

As $\rho \to 1$, the denominator $1 - \rho$ becomes small. Economically, if trade signs are nearly perfectly persistent, the unexpected component $x_t - \rho x_{t-1}$ has little independent variation. It becomes difficult to separate a genuinely informative innovation in order flow from a predictable continuation of the same trading program.

The coefficient θ measures permanent price impact from unexpected order flow. The coefficient ϕ measures the transitory component, often associated with order-processing costs, immediacy, inventory compensation, or bid-ask bounce. □

Takeaway. *MRR identification is a regression anatomy lesson. The current trade coefficient is not pure information, and the lagged trade coefficient is not pure reversal; only their structure, together with trade-sign persistence, separates permanent and transitory effects.*

References for this problem.

- Madhavan, Richardson, and Roomans (1997), for identifying permanent and transitory components from transaction-level regressions.
- Hasbrouck (2007), for empirical price-impact regression methods.
- Glosten and Harris (1988), for a related spread decomposition using trade size.

1.5 Spread Decomposition and Multi-Market Price Discovery

Problem 1.8 (Advanced: Glosten-Harris size-based spread decomposition)**.** Let $Q_t \in \{-1, +1\}$ denote trade sign and let $V_t > 0$ denote trade size. Suppose the permanent adverse-selection component and the transitory order-processing component are

$$Z_t = z_0 + z_1 V_t, \qquad C_t = c_0 + c_1 V_t.$$

Consider the simplified transaction-price change equation

$$\Delta p_t = \epsilon_t + Q_t Z_t + Q_t C_t - Q_{t-1} C_{t-1}.$$

(a) Identify which terms are permanent and which terms are transitory.

(b) Explain how variation in V_t can help distinguish adverse selection from order-processing costs.

(c) What does $z_1 > 0$ mean economically?

(d) What does $c_1 > 0$ mean economically?

(e) Describe an empirical regression that could estimate these components.

Solution.

The term $Q_t Z_t$ is permanent because it enters the efficient price. A buy with a large permanent component raises the assessed value; a sell lowers it. The term

$$Q_t C_t - Q_{t-1} C_{t-1}$$

is transitory because the current transaction-cost component enters today's price but the previous transaction-cost component reverses out of the price change.

Trade size helps because information and execution cost need not scale in the same way. If large trades are more likely to be privately informed, then the size slope in the permanent component, z_1, should be positive. If larger trades consume more liquidity or require more compensation for immediacy, then the size slope in the transitory component, c_1, should be positive.

Substituting the component definitions gives

$$\begin{aligned}\Delta p_t = \epsilon_t + z_0 Q_t + z_1 Q_t V_t + c_0 Q_t + c_1 Q_t V_t \\ - c_0 Q_{t-1} - c_1 Q_{t-1} V_{t-1}.\end{aligned}$$

In a linear regression of transaction-price changes on Q_t, $Q_t V_t$, Q_{t-1}, and $Q_{t-1} V_{t-1}$, the current signed-size terms combine permanent and transitory effects, while the lagged signed-size terms identify the reversal in the transitory component. The identifying logic is that adverse selection persists but order-processing cost reverses. □

Takeaway. *Glosten-Harris turns the spread from one number into a decomposition. Size is useful because permanent price impact and temporary execution cost leave different dynamic footprints in transaction returns.*

References for this problem.

- Glosten and Harris (1988), for fixed and size-dependent spread components.
- Hasbrouck (2007), for empirical transaction-cost measurement.
- Madhavan, Richardson, and Roomans (1997), for the related permanent/transitory decomposition.

Problem 1.9 (Advanced: Hasbrouck information share in two markets)**.** Two markets trade claims on the same asset. Let their price innovations in a cointegrated vector error-correction model be

$$e_t = (e_{1t}, e_{2t})', \qquad \text{Var}(e_t) = \Omega.$$

Suppose the innovation in the common efficient price is $w' e_t$, where w is the common-trend loading vector.

(a) Show that the variance of the common efficient-price innovation is $w'\Omega w$.

(b) Let $\Omega = FF'$ be a Cholesky factorization and write $e_t = Fz_t$, where $\text{Var}(z_t) = I$. Express $w'e_t$ in terms of z_t.

(c) Define the information share of each market under this ordering.

(d) Why does the Cholesky ordering matter?

(e) How are upper and lower information-share bounds computed?

Solution.

Since the efficient-price innovation is $w'e_t$,

$$\text{Var}(w'e_t) = w' \, \text{Var}(e_t) w = w'\Omega w.$$

If $\Omega = FF'$ and $e_t = Fz_t$, then

$$w'e_t = w'Fz_t.$$

Let

$$\ell' = w'F.$$

Then

$$w'e_t = \ell_1 z_{1t} + \ell_2 z_{2t},$$

and because the components of z_t are orthogonal with unit variance,

$$\text{Var}(w'e_t) = \ell_1^2 + \ell_2^2 = w'\Omega w.$$

Under this ordering, market j's information share is

$$IS_j = \frac{\ell_j^2}{w'\Omega w}.$$

The Cholesky ordering matters because contemporaneously correlated innovations must be orthogonalized before assigning variance shares. The market placed first receives the common contemporaneous component first; the market placed last receives only the residual component after the earlier market's innovation has been accounted for. Therefore empirical work reports bounds: place a market first to obtain its upper bound and last to obtain its lower bound. □

Takeaway. *Hasbrouck information share measures contribution to the permanent common-price innovation. It is a price-discovery object, not a direct measure of trading cost, depth, or execution quality.*

References for this problem.

- Hasbrouck (1995), for information shares in a cointegrated multi-market system.
- Hasbrouck (2007), for the textbook treatment of price discovery and vector error-correction models.
- Gonzalo and Granger (1995), for a related common-factor decomposition.

Problem 1.10 (Core: When information share is not execution quality)**.** Construct a two-market example in which Market A has almost all of the Hasbrouck information share but Market B offers lower trading costs. Explain why this is not a contradiction.

(a) Give a plausible microstructure story.

(b) State what information share measures.

(c) State what spread, depth, or execution-cost measures capture.

(d) Explain why informed trading can raise both price discovery and trading costs in the same venue.

Solution.

Suppose informed traders prefer Market A because it is faster, has colocated participants, or reacts first to public and private signals. Then innovations in Market A arrive first and Market B follows. In a common-trend decomposition, Market A will explain most of the permanent efficient-price innovation.

At the same time, liquidity suppliers in Market A understand that order flow there is more toxic. They protect themselves by quoting wider spreads or reducing displayed depth. Market B may be slower but less exposed to informed flow, so it can offer narrower displayed spreads for small uninformed trades.

There is no contradiction because the two measurements answer different questions. Hasbrouck information share asks: where is the common efficient price impounded first? Execution-quality measures ask: where can a trader execute cheaply, with depth and low price impact? The informed venue can dominate price discovery precisely because it faces more adverse selection. □

Takeaway. *A market can be important for price discovery and unattractive for cheap execution. Graduate students should not read "moves first" as "best market" without looking at trading costs and liquidity supply.*

References for this problem.

- Hasbrouck (1995), for the distinction between permanent price discovery and market location.

- Hasbrouck (2007), for information shares and market quality metrics.
- Gonzalo and Granger (1995), for an alternative permanent-transitory decomposition.

Problem 1.11 (Research: Reconciling spread, impact, and information share)**.** After a tick-size reduction or venue technology upgrade, a researcher estimates three facts:

Roll spread falls,
MRR permanent impact rises,
Hasbrouck information share shifts from A to B.

Give a coherent economic explanation for all three findings.

(a) Explain why the Roll spread can fall after a tick-size reduction.

(b) Explain why MRR permanent impact can rise at the same time.

(c) Explain why information share can migrate across venues.

(d) What additional empirical evidence would you want before calling the market more efficient?

Solution.

A tick-size reduction can reduce the mechanical bid-ask bounce component of transaction prices. If quoted and effective spreads compress, the Roll autocovariance-based spread estimate can fall because the transitory component of transaction returns is smaller.

At the same time, informed traders may concentrate in the faster or more competitive venue. If order flow becomes more information-rich, the MRR permanent-impact coefficient can rise even though transitory trading costs have fallen. Lower explicit trading costs can make informed trading more aggressive, so permanent impact and mechanical spread need not move in the same direction.

Hasbrouck information share can shift from A to B if B begins incorporating common efficient-price innovations first. A technology upgrade, better routing, lower latency, or a change in participant composition can make B the price discovery venue.

Before calling the market more efficient, one would want evidence on short-horizon return predictability, quote adjustment speed, depth, realized spreads, adverse-selection costs, cancellation intensity, and execution quality for different trader types. A lower Roll spread alone is not enough; a higher permanent impact coefficient may indicate more information production or more severe adverse selection. □

Takeaway. *Market-structure changes move several margins at once. A serious empirical interpretation must separate transitory cost, permanent information impact, and the venue in which common-price innovations appear first.*

References for this problem.

- Roll (1984), for the bid-ask spread moment restriction.
- Madhavan, Richardson, and Roomans (1997), for permanent and transitory price impact.
- Hasbrouck (1995), for information shares across markets.
- Glosten and Harris (1988), for trade-size spread decomposition.

1.6 Price Histories, Invertibility, and Stale Prices

Problem 1.12 (Advanced: Wold invertibility of transaction returns)**.** Return to the Roll model

$$p_t = m_t + cq_t, \qquad m_t = m_{t-1} + u_t,$$

where u_t is iid with variance σ_u^2, q_t is iid with mean zero and unit variance, and u_t is independent of q_s for all s. Let $r_t = \Delta p_t$.

(a) Show that r_t has only first-order serial dependence.

(b) Match the autocovariances of r_t to an invertible MA(1) representation

$$r_t = \eta_t + \vartheta\eta_{t-1}, \qquad |\vartheta| < 1, \qquad \mathrm{Var}(\eta_t) = \sigma_\eta^2.$$

(c) Derive the equation that determines ϑ from γ_1/γ_0.

(d) Explain why an invertible Wold representation of returns is not the same thing as observing the efficient-price innovation u_t.

Solution.

From Roll,

$$r_t = u_t + c(q_t - q_{t-1}).$$

Therefore

$$\gamma_0 = \mathrm{Var}(r_t) = \sigma_u^2 + 2c^2, \qquad \gamma_1 = \mathrm{Cov}(r_t, r_{t-1}) = -c^2,$$

and $\gamma_k = 0$ for $|k| > 1$. Thus r_t has the autocovariance pattern of an MA(1).

For the invertible MA(1),

$$\gamma_0 = (1 + \vartheta^2)\sigma_\eta^2, \qquad \gamma_1 = \vartheta\sigma_\eta^2.$$

Taking the ratio gives

$$\frac{\gamma_1}{\gamma_0} = \frac{\vartheta}{1 + \vartheta^2}.$$

Since $\gamma_1/\gamma_0 < 0$, the invertible root has $\vartheta < 0$ and $|\vartheta| < 1$. Once ϑ is found,

$$\sigma_\eta^2 = \frac{\gamma_1}{\vartheta}.$$

The Wold innovation η_t is the one-step-ahead forecast error from the history of transaction returns. It is not generally equal to the efficient-price innovation u_t. Transaction returns mix efficient-price news with bid-ask bounce. Without trade signs or additional structure, the price history alone reveals an innovation in observed returns, not a clean decomposition into fundamental news and trading-cost noise. □

Takeaway. *Invertibility is a statement about recovering return shocks from return histories. It is not automatically a statement about recovering the latent efficient-price innovation that a market microstructure model cares about.*

References for this problem.

- Roll (1984), for bid-ask bounce as a transitory pricing-error model.
- Hamilton (1994), for Wold representation and invertibility.
- Hasbrouck (2007), for linear time-series tools in transaction-price econometrics.

Problem 1.13 (Core: Efficient-price projection from transaction returns)**.** In the Roll model above, suppose the econometrician observes only the transaction return $r_t = \Delta p_t$ and wants to estimate the efficient-price innovation u_t linearly from r_t.

(a) Compute the best linear predictor $\widehat{u}_t = \beta r_t$.

(b) Compute the projection error variance $\mathrm{Var}(u_t - \widehat{u}_t)$.

(c) Study the limiting cases $c \to 0$ and $c \to \infty$.

(d) Interpret the result as a signal-to-noise problem.

Solution.

The best linear coefficient is

$$\beta = \frac{\mathrm{Cov}(u_t, r_t)}{\mathrm{Var}(r_t)}.$$

Since

$$r_t = u_t + c(q_t - q_{t-1}),$$

and u_t is independent of trade signs,

$$\mathrm{Cov}(u_t, r_t) = \sigma_u^2, \qquad \mathrm{Var}(r_t) = \sigma_u^2 + 2c^2.$$

Therefore

$$\widehat{u}_t = \frac{\sigma_u^2}{\sigma_u^2 + 2c^2} r_t.$$

The projection error variance is

$$\begin{aligned} \operatorname{Var}(u_t - \widehat{u}_t) &= \operatorname{Var}(u_t) - \frac{\operatorname{Cov}(u_t, r_t)^2}{\operatorname{Var}(r_t)} \\ &= \sigma_u^2 - \frac{\sigma_u^4}{\sigma_u^2 + 2c^2} \\ &= \frac{2c^2 \sigma_u^2}{\sigma_u^2 + 2c^2}. \end{aligned}$$

If $c \to 0$, then transaction returns equal efficient-price innovations and the error variance goes to zero. If $c \to \infty$, transaction returns are dominated by bid-ask bounce, $\beta \to 0$, and the error variance tends to σ_u^2: observing the transaction return becomes nearly useless for inferring u_t. □

Takeaway. *The efficient-price innovation is a signal hidden inside transaction noise. Large spreads reduce the informational content of raw transaction returns even when the underlying efficient price is a simple random walk.*

References for this problem.

- Roll (1984), for the transitory pricing-error structure.
- Hasbrouck (2007), for efficient-price filtering and transaction-price representations.
- Hamilton (1994), for linear projections and time-series filtering.
- Madhavan, Richardson, and Roomans (1997), for permanent impact as an efficient-price component.

Problem 1.14 (Core: Stale-price dynamics and delayed adjustment)**.** Let the efficient price follow

$$m_t = m_{t-1} + u_t, \qquad \operatorname{Var}(u_t) = \sigma_u^2.$$

Observed prices adjust only partially within the period:

$$p_t = (1 - \delta) m_t + \delta m_{t-1}, \qquad 0 \le \delta \le 1.$$

(a) Derive $r_t = \Delta p_t$ in terms of efficient-price innovations.

(b) Compute $\operatorname{Var}(r_t)$ and $\operatorname{Cov}(r_t, r_{t-1})$.

(c) Explain why stale prices create positive rather than negative return autocovariance.

(d) How does this differ from bid-ask bounce?

Solution.

Since

$$p_t = (1-\delta)m_t + \delta m_{t-1} = m_{t-1} + (1-\delta)u_t,$$

and

$$p_{t-1} = m_{t-2} + (1-\delta)u_{t-1},$$

we have

$$r_t = p_t - p_{t-1} = u_{t-1} + (1-\delta)u_t - (1-\delta)u_{t-1}.$$

Thus

$$r_t = (1-\delta)u_t + \delta u_{t-1}.$$

The variance is

$$\mathrm{Var}(r_t) = \left\{(1-\delta)^2 + \delta^2\right\}\sigma_u^2.$$

The first autocovariance is

$$\mathrm{Cov}(r_t, r_{t-1}) = \mathrm{Cov}((1-\delta)u_t + \delta u_{t-1}, (1-\delta)u_{t-1} + \delta u_{t-2}) = \delta(1-\delta)\sigma_u^2.$$

This is positive for $0 < \delta < 1$.

The intuition is that stale prices spread the same efficient-price innovation across adjacent observed returns. If only part of today's efficient-price move is reflected today, the remainder appears tomorrow. Bid-ask bounce does the opposite: it mechanically reverses a transitory trading-cost component and therefore creates negative autocovariance. □

Takeaway. *Not all serial dependence in high-frequency returns is bid-ask bounce. Stale prices and delayed adjustment create positive autocorrelation because news is incorporated gradually.*

References for this problem.

- Scholes and Williams (1977), for nonsynchronous-trading beta bias.
- Dimson (1979), for infrequent-trading beta adjustment.
- Lo and MacKinlay (1990), for the econometric analysis of nonsynchronous trading.
- Hasbrouck (2007), for microstructure treatment of stale prices.

Problem 1.15 (Advanced: Correlated delays in two securities)**.** Two securities have efficient prices

$$m_{i,t} = m_{i,t-1} + u_{i,t}, \qquad i = 1, 2,$$

with

$$\mathrm{Cov}(u_{1,t}, u_{2,t}) = \sigma_{12}, \qquad \mathrm{Cov}(u_{1,t}, u_{2,s}) = 0 \quad \text{for } t \neq s.$$

Observed prices adjust with possibly different delays:

$$p_{i,t} = (1-\delta_i)m_{i,t} + \delta_i m_{i,t-1}.$$

(a) Derive the observed return $r_{i,t}$.

(b) Compute $\text{Cov}(r_{1,t}, r_{2,t})$.

(c) Compute $\text{Cov}(r_{1,t}, r_{2,t-1})$ and $\text{Cov}(r_{1,t-1}, r_{2,t})$.

(d) Explain how nonsynchronous adjustment can generate lead-lag effects even when efficient-price innovations are contemporaneously correlated only.

Solution.

From the previous problem,

$$r_{i,t} = (1-\delta_i)u_{i,t} + \delta_i u_{i,t-1}.$$

The contemporaneous observed covariance is

$$\begin{aligned}\text{Cov}(r_{1,t}, r_{2,t}) &= (1-\delta_1)(1-\delta_2)\sigma_{12} + \delta_1\delta_2\sigma_{12} \\ &= \{(1-\delta_1)(1-\delta_2) + \delta_1\delta_2\}\,\sigma_{12}.\end{aligned}$$

The cross-lag covariance is

$$\begin{aligned}\text{Cov}(r_{1,t}, r_{2,t-1}) &= \text{Cov}((1-\delta_1)u_{1,t} + \delta_1 u_{1,t-1}, (1-\delta_2)u_{2,t-1} + \delta_2 u_{2,t-2}) \\ &= \delta_1(1-\delta_2)\sigma_{12}.\end{aligned}$$

Similarly,

$$\text{Cov}(r_{1,t-1}, r_{2,t}) = (1-\delta_1)\delta_2\sigma_{12}.$$

If $\delta_1 > \delta_2$, security 1 adjusts more slowly, so news shared by both efficient prices appears later in security 1's observed returns. This produces lead-lag covariance in observed prices even though the efficient innovations have no true lagged cross-covariance. □

Takeaway. *Lead-lag effects in high-frequency returns can reflect trading frictions and delayed quote adjustment, not necessarily delayed information production in the underlying fundamentals.*

References for this problem.

- Lo and MacKinlay (1990), for nonsynchronous trading and induced lead-lag covariance.
- Scholes and Williams (1977), for the original nonsynchronous beta correction.
- Dimson (1979), for infrequent-trading adjustment.

- Hasbrouck (2007), for price-discovery implications of asynchronous trading.

Problem 1.16 (Research: Simulation lab for spread and delay estimators)**.** Design a simulation that compares three data-generating processes:

Roll bid-ask bounce, MRR permanent impact, stale-price delayed adjustment.

(a) Specify a parameter vector for each model.

(b) State which moments you would estimate from each simulated sample.

(c) Explain which estimator should work well in each model and which should fail.

(d) Give pseudocode for the simulation.

Solution.

A clean design uses the following parameter blocks:

$$(\sigma_u, c) \quad \text{for Roll}, \qquad (\sigma_u, c, \lambda, \rho) \quad \text{for MRR}, \qquad (\sigma_u, \delta) \quad \text{for stale prices}.$$

For each simulated path, estimate

$$\widehat{\gamma}_0, \quad \widehat{\gamma}_1, \quad \widehat{\gamma}_2,$$

and, when trade signs are observed, regress Δp_t on q_t and q_{t-1}.

The Roll estimator should work well in the pure Roll model with iid signs and large samples. It should fail or become biased in the MRR model because $\gamma_1 = -c(\lambda + c)$ rather than $-c^2$. It should also fail in the stale-price model because stale prices generate positive autocovariance, while Roll expects negative autocovariance. The MRR regression should recover permanent and transitory components when the MRR specification is correct and trade signs are measured correctly.

One possible pseudocode block is:

1. Choose T, N and parameter grids.
2. For each replication $n = 1, \ldots, N$:
3. simulate efficient innovations u_t.
4. simulate trade signs q_t, iid or Markov.
5. construct p_t under Roll, MRR, or stale adjustment.
6. estimate autocovariances and regression coefficients.
7. store bias, RMSE, undefined Roll estimates, and coverage.
8. Summarize performance across models and parameter values.

□

Takeaway. *A simulation lab turns identification assumptions into visible failures. The same return autocovariance can mean bid-ask bounce, adverse selection, or delayed adjustment depending on the data-generating process.*

References for this problem.

- Roll (1984), for the bid-ask bounce estimator.
- Madhavan, Richardson, and Roomans (1997), for permanent/transitory impact.
- Lo and MacKinlay (1990), for stale-price and nonsynchronous-trading distortions.
- Harris (1990), for finite-sample behavior of the Roll estimator.

1.7 Modern Noise, Synchronization, and High-Dimensional Price Discovery

The final block of the foundations chapter connects the classic transaction price models to recent high-frequency econometrics. The common theme is that microstructure noise is no longer merely an iid error. It can depend on order book state, trade intensity, clock time, and the asynchronous way assets trade.

Problem 1.17 (Research: Endogenous microstructure noise from trading information)**.** Let the observed log price be

$$p_{t_i} = m_{t_i} + \eta_{t_i},$$

where m is the latent efficient price and the microstructure noise satisfies

$$\eta_{t_i} = g(X_{t_i}) + \varepsilon_{t_i}.$$

Here X_{t_i} is a vector of trading information such as the bid-ask spread, depth imbalance, quoted depth, and trade intensity. Assume $\mathbb{E}[\varepsilon_{t_i} \mid X_{t_i}] = 0$ and, for simplicity, ε_{t_i} is serially uncorrelated with variance ω^2.

(a) Derive the observed return Δp_{t_i}.

(b) Show why realized variance computed from Δp_{t_i} estimates efficient variation plus noise variation.

(c) Suppose $\widehat{g}(X_{t_i})$ is a nonparametric estimate of $g(X_{t_i})$. Construct a residualized price and explain how it reduces noise bias.

(d) Explain why ignoring X_{t_i} can make the noise look serially dependent even when ε_{t_i} is not.

Solution.

The observed return is

$$\Delta p_{t_i} = \Delta m_{t_i} + \{g(X_{t_i}) - g(X_{t_{i-1}})\} + \Delta\varepsilon_{t_i}.$$

The realized variance is

$$RV_p = \sum_i (\Delta p_{t_i})^2.$$

Expanding term by term gives

$$\begin{aligned} RV_p &= \sum_i (\Delta m_{t_i})^2 + \sum_i \{\Delta g(X_{t_i})\}^2 + \sum_i (\Delta\varepsilon_{t_i})^2 \\ &\quad + 2\sum_i \Delta m_{t_i}\Delta g(X_{t_i}) + 2\sum_i \Delta m_{t_i}\Delta\varepsilon_{t_i} + 2\sum_i \Delta g(X_{t_i})\Delta\varepsilon_{t_i}. \end{aligned}$$

Even if the cross terms are mean zero, the second and third sums add positive variation. In particular, if ε_{t_i} is iid with variance ω^2, then

$$\mathbb{E}[(\Delta\varepsilon_{t_i})^2] = 2\omega^2,$$

so the noise contribution grows with the number of observations.

A residualized price is

$$\widetilde{p}_{t_i} = p_{t_i} - \widehat{g}(X_{t_i}).$$

Then

$$\Delta\widetilde{p}_{t_i} = \Delta m_{t_i} + \Delta\{g(X_{t_i}) - \widehat{g}(X_{t_i})\} + \Delta\varepsilon_{t_i}.$$

If $\widehat{g}$ is accurate, the systematic trading-information component is removed and the remaining noise is closer to a residual nuisance.

Ignoring X_{t_i} can create apparent serial dependence because spreads, depth, and intensity are persistent. If $g(X_{t_i})$ changes slowly, then $\Delta g(X_{t_i})$ has its own autocorrelation pattern. The econometrician who estimates only $p = m + \eta$ may attribute this predictable component to dependent noise, even though the residual ε is conditionally uncorrelated. □

Takeaway. *Modern microstructure noise is partly explainable. Treating it as iid hides the fact that order-book variables can be used to remove systematic measurement error from high-frequency prices.*

References for this problem.

- Cui, Hu, and Wang (2024), for nonparametric estimation with trading information.
- Ait-Sahalia and Jacod (2014), for high-frequency econometrics with market microstructure noise.

- Zhang, Mykland, and Ait-Sahalia (2005), for two-time-scale volatility correction under noise.

Problem 1.18 (Advanced: Pre-averaging under serially dependent bid-ask noise)**.** Let observations arrive on an equally spaced grid and suppose

$$p_i = m_i + \eta_i, \qquad i = 0, \dots, n,$$

where m_i is an efficient price sampled from a continuous semimartingale. Define a pre-averaged return over a window of length k by

$$\bar{r}_i = \sum_{j=1}^{k-1} w_j (p_{i+j} - p_{i+j-1}), \qquad w_j = \frac{j}{k}\left(1 - \frac{j}{k}\right).$$

(a) Decompose $\bar{r}_i$ into efficient-price and noise components.

(b) Explain why pre-averaging reduces the influence of high-frequency noise relative to raw one-tick returns.

(c) Show the bias-variance tradeoff in choosing k.

(d) Explain why serially dependent noise requires an explicit finite-sample correction or a robust asymptotic argument.

Solution.

Using $p_i = m_i + \eta_i$,

$$\bar{r}_i = \sum_{j=1}^{k-1} w_j \Delta m_{i+j} + \sum_{j=1}^{k-1} w_j \Delta \eta_{i+j}.$$

The first term is a local average of efficient returns. The second term is a weighted average of noise differences.

Raw one-tick returns contain $\Delta\eta_i$, which is often large relative to the efficient return at very high frequency. Pre-averaging smooths adjacent returns, so independent or weakly dependent noise partially cancels. The efficient component is also averaged, but because efficient variation accumulates over time while noise oscillates at the observation scale, the signal-to-noise ratio can improve.

The window length k governs the tradeoff. If k is too small, the statistic looks like raw realized variance and remains dominated by microstructure noise. If k is too large, the estimator oversmooths local price variation and loses temporal resolution. In high-frequency asymptotics, k is usually allowed to grow with n but not as fast as n.

If η_i is serially dependent, then

$$\text{Var}\left(\sum_{j=1}^{k-1} w_j \Delta\eta_{i+j}\right)$$

depends on the autocovariances of η_i, not only on $\text{Var}(\eta_i)$. A correction calibrated for iid noise will generally be wrong. Robust pre-averaging or kernel estimators must account for the local dependence structure in the noise process. □

Takeaway. *Pre-averaging is the volatility analogue of looking past bid-ask bounce. The statistical difficulty is choosing enough smoothing to kill noise without washing away the efficient price.*

References for this problem.

- Jacod et al. (2009), for the pre-averaging approach to microstructure noise.
- Christensen, Kinnebrock, and Podolskij (2010), for pre-averaged covariance with noisy nonsynchronous data.
- Ait-Sahalia and Jacod (2014), for the textbook treatment of noise-robust volatility estimation.
- Zhang, Mykland, and Ait-Sahalia (2005), for two-time-scale estimation.

Problem 1.19 (Research: Asynchronous high-dimensional price discovery and the Epps effect)**.** Let two latent efficient prices satisfy

$$\Delta m_{1,t} = u_{1,t}, \qquad \Delta m_{2,t} = u_{2,t}, \qquad \text{Cov}(u_{1,t}, u_{2,t}) = \sigma_{12}.$$

Asset 1 is observed every period, while asset 2 is observed only with probability π in each period. When asset 2 is not observed, its last transaction price is carried forward. Ignore bid-ask bounce.

(a) Write observed returns for asset 2 when it trades after L silent periods.

(b) Explain why contemporaneous sampled covariance understates σ_{12} at very high frequency.

(c) Define a refresh-time return pair for the two assets.

(d) Explain how the same idea extends to many assets and why high dimensionality creates a new estimation problem.

Solution.

If asset 2 last traded at time $t - L - 1$ and then trades at time t, its observed return cumulates all latent innovations since the previous trade:

$$r_{2,t}^{obs} = \sum_{\ell=0}^{L} u_{2,t-\ell}.$$

At periods with no trade, the carried-forward price produces observed return zero.

The contemporaneous sampled covariance is biased downward because many efficient innovations in asset 2 are not recorded at the same clock time as the corresponding asset 1 innovations. Some covariance appears at leads and lags instead of at lag zero. This is the Epps effect: measured correlation falls as the sampling interval becomes too fine relative to trading intensity.

Refresh time solves this by waiting until both assets have traded since the last refresh. The paired return is the change in each asset's most recent transaction price between consecutive refresh times. This aligns information sets by event availability rather than by a rigid clock.

In many assets, refresh time can become sparse because one must wait for all assets to update. High-dimensional estimators therefore combine synchronization ideas with shrinkage, factor structure, or low-rank plus sparse covariance assumptions. The econometric problem is no longer just correcting one noisy covariance; it is estimating a stable covariance matrix from asynchronous, noisy, irregularly observed prices. □

Takeaway. *At high frequency, correlation can disappear because clocks are wrong, not because economics disappeared. Synchronization is part of price discovery.*

References for this problem.

- Lo and MacKinlay (1990), for nonsynchronous trading and covariance distortion.
- Hayashi and Yoshida (2005), for covariance estimation with nonsynchronous observations.
- Ait-Sahalia and Jacod (2014), for high-frequency covariance estimation under noise and nonsynchronicity.
- Christensen, Kinnebrock, and Podolskij (2010), for pre-averaged covariance with noisy nonsynchronous data.

Bibliography

Ait-Sahalia, Yacine, and Jean Jacod (2014). *High-Frequency Financial Econometrics.* Princeton University Press. [Textbook source for noise-robust volatility and covariance estimation.]

Christensen, Kim, Silja Kinnebrock, and Mark Podolskij (2010). “Pre-averaging Estimators of the Ex-post Covariance Matrix in Noisy Diffusion Models with Non-synchronous Data.” *Journal of Econometrics* 159(1), 116–133. [Primary extension for pre-averaged covariance under noise and nonsynchronicity.]

Cui, Wenhao, Jie Hu, and Jiandong Wang (2024). “Nonparametric Estimation for High-Frequency Data Incorporating Trading Information.” *Journal of Econometrics* 240(1), 105690. [Modern source for microstructure noise depending on trading information.]

Dimson, Elroy (1979). “Risk Measurement When Shares Are Subject to Infrequent Trading.” *Journal of Financial Economics* 7(2), 197–226. [Canonical correction for infrequent-trading/stale-price bias.]

Glosten, Lawrence R., and Lawrence E. Harris (1988). “Estimating the Components of the Bid/Ask Spread.” *Journal of Financial Economics* 21(1), 123–142. [Primary size-based spread decomposition source.]

Glosten, Lawrence R., and Paul R. Milgrom (1985). “Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders.” *Journal of Financial Economics* 14(1), 71–100. [Primary Bayesian adverse-selection quote-setting source.]

Gonzalo, Jesus, and Clive Granger (1995). “Estimation of Common Long-Memory Components in Cointegrated Systems.” *Journal of Business & Economic Statistics* 13(1), 27–35. [Common-factor comparison for Hasbrouck information shares.]

Hamilton, James D. (1994). *Time Series Analysis.* Princeton University Press. [Textbook source for Wold representation and invertibility.]

Harris, Lawrence (1990). “Statistical Properties of the Roll Serial Covariance Bid/Ask Spread Estimator.” *Journal of Finance* 45(2), 579–590. [Primary finite-sample Roll estimator source.]

Hasbrouck, Joel (1995). “One Security, Many Markets: Determining the Contributions to Price Discovery.” *Journal of Finance* 50(4), 1175–1199. [Primary information-share source.]

Hasbrouck, Joel (2007). *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading.* Oxford University Press. [Main first-reading textbook for empirical microstructure.]

Hayashi, Takaki, and Nakahiro Yoshida (2005). “On Covariance Estimation of Non-synchronously Observed Diffusion Processes.” *Bernoulli* 11(2), 359–379. [Primary asynchronous covariance estimation source.]

Jacod, Jean, Yingying Li, Per A. Mykland, Mark Podolskij, and Mathias Vetter (2009). "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach." *Stochastic Processes and their Applications* 119(7), 2249–2276. [Primary pre-averaging source.]

Lee, Charles M. C., and Mark J. Ready (1991). "Inferring Trade Direction from Intraday Data." *Journal of Finance* 46(2), 733–746. [Canonical trade-sign classification source.]

Lo, Andrew W., and A. Craig MacKinlay (1990). "An Econometric Analysis of Nonsynchronous Trading." *Journal of Econometrics* 45(1–2), 181–211. [Primary econometric source for nonsynchronous/stale trading.]

Madhavan, Ananth, Matthew Richardson, and Mark Roomans (1997). "Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks." *Review of Financial Studies* 10(4), 1035–1064. [Primary MRR permanent/transitory price-impact source.]

O'Hara, Maureen (1995). *Market Microstructure Theory*. Blackwell. [Main theory textbook for adverse selection and sequential trade.]

Roll, Richard (1984). "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market." *Journal of Finance* 39(4), 1127–1139. [Primary Roll spread identification source.]

Scholes, Myron, and Joseph Williams (1977). "Estimating Betas from Nonsynchronous Data." *Journal of Financial Economics* 5(3), 309–327. [Canonical stale-price/nonsynchronous-trading source.]

Zhang, Lan, Per A. Mykland, and Yacine Ait-Sahalia (2005). "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data." *Journal of the American Statistical Association* 100(472), 1394–1411. [Two-time-scale volatility correction under microstructure noise.]

Chapter 2

Sequential Trade and Adverse Selection Models

The first chapter treated order flow mostly through moments, regressions, and linear decompositions. This chapter changes the lens. Order flow is now an information process: trades, trade sizes, and even no-trade intervals move beliefs about a latent asset value.

2.1 Bayesian Quotes and Sequential Learning

Problem 2.1 (Core: Glosten-Milgrom posterior bid and ask quotes)**.** Let the asset value be binary,

$$V \in \{v_L, v_H\}, \qquad \mathbb{P}(V = v_H) = p,$$

where $v_H > v_L$. A trader is informed with probability μ and uninformed with probability $1 - \mu$. Informed traders buy when $V = v_H$ and sell when $V = v_L$. Uninformed traders buy and sell with equal probability.

(a) Compute $\mathbb{P}(v_H \mid B)$ after a buy order.

(b) Compute $\mathbb{P}(v_H \mid S)$ after a sell order.

(c) Derive the competitive ask $a = \mathbb{E}[V \mid B]$ and bid $b = \mathbb{E}[V \mid S]$.

(d) Show that the spread is zero when $\mu = 0$.

(e) Explain how the spread changes with private-information intensity and prior uncertainty.

Solution.

The probability of a buy conditional on the high state is

$$\mathbb{P}(B \mid v_H) = \mu + \frac{1-\mu}{2} = \frac{1+\mu}{2},$$

while

$$\mathbb{P}(B \mid v_L) = \frac{1-\mu}{2}.$$

Bayes' rule gives

$$\mathbb{P}(v_H \mid B) = \frac{p(1+\mu)}{p(1+\mu) + (1-p)(1-\mu)}.$$

Similarly,

$$\mathbb{P}(v_H \mid S) = \frac{p(1-\mu)}{p(1-\mu) + (1-p)(1+\mu)}.$$

The competitive ask and bid are conditional expectations:

$$a = v_L + (v_H - v_L)\mathbb{P}(v_H \mid B),$$

and

$$b = v_L + (v_H - v_L)\mathbb{P}(v_H \mid S).$$

Therefore

$$a - b = (v_H - v_L)\{\mathbb{P}(v_H \mid B) - \mathbb{P}(v_H \mid S)\}.$$

If $\mu = 0$, then buys and sells are equally likely in both states, so

$$\mathbb{P}(v_H \mid B) = \mathbb{P}(v_H \mid S) = p$$

and the spread is zero.

As μ rises, a buy becomes stronger evidence of the high state and a sell becomes stronger evidence of the low state, so the spread widens. The spread is also largest when prior uncertainty is high, because order flow is then more informative about value. □

Takeaway. *The bid-ask spread can arise even with risk-neutral competitive dealers and no inventory cost. The spread compensates the dealer for the option given to informed traders.*

References for this problem.

- Glosten and Milgrom (1985), for Bayesian bid and ask quotes under adverse selection.
- O'Hara (1995), for the textbook treatment of sequential trade and market-maker learning.
- Easley and O'Hara (1987), for trade size as an additional information signal.

Problem 2.2 (Core: Sequential filtering from buy and sell counts)**.** Use the binary-value model from the previous problem. Suppose that by time t the market maker has observed n_B buys and

n_S sells.

(a) Derive the posterior odds ratio
$$\frac{\mathbb{P}(v_H \mid \mathcal{F}_t)}{\mathbb{P}(v_L \mid \mathcal{F}_t)}.$$

(b) Show how a run of buys changes the next ask.

(c) Explain why price impact is state-dependent in this model.

(d) Compare this nonlinear filtering view with a linear MRR impact coefficient.

Solution.

Conditional on the high state, a buy has likelihood $(1 + \mu)/2$ and a sell has likelihood $(1 - \mu)/2$. Conditional on the low state, the likelihoods are reversed. Hence the posterior odds are

$$\frac{\mathbb{P}(v_H \mid \mathcal{F}_t)}{\mathbb{P}(v_L \mid \mathcal{F}_t)} = \frac{p}{1-p}\left(\frac{1+\mu}{1-\mu}\right)^{n_B - n_S}.$$

Only the order imbalance $n_B - n_S$ matters for the posterior in this simple model.

After a run of buys, the posterior probability of the high state rises. The next ask,

$$a_t = \mathbb{E}[V \mid \mathcal{F}_{t-}, B],$$

also rises. But the marginal effect of another buy is not constant. When the prior is near 1/2, a buy can move beliefs substantially. When the posterior is already close to one, another buy has little room to increase the expected value.

MRR summarizes permanent price impact with an average linear coefficient. The sequential Bayesian model shows what that coefficient is averaging over: different belief states, different prior uncertainty, and different histories of order imbalance. □

Takeaway. *Bayesian market microstructure turns order imbalance into a sufficient statistic for beliefs. Price impact is naturally nonlinear because the same trade has a different meaning at different posterior beliefs.*

References for this problem.

- Glosten and Milgrom (1985), for Bayesian quote updating from order flow.
- O'Hara (1995), for sequential-trade exposition.
- Madhavan, Richardson, and Roomans (1997), for the reduced-form average of structural belief updating.

Problem 2.3 (Advanced: Trade size as an information signal)**.** Extend the Glosten-Milgrom model by allowing orders to be either small or large. Informed traders choose a large order with probability θ_I, while uninformed traders choose a large order with probability θ_U, where

$$\theta_I > \theta_U.$$

Let Lg denote a large order and Sm a small order.

(a) Derive a formula for $\mathbb{P}(v_H \mid B, Lg)$.

(b) Derive a formula for $\mathbb{P}(v_H \mid B, Sm)$.

(c) Show that large buys have higher adverse-selection content than small buys.

(d) Propose a regression that tests whether large trades have larger permanent price impact.

Solution.

The likelihood of observing a large buy in the high state is

$$\mathbb{P}(B, Lg \mid v_H) = \mu\theta_I + (1-\mu)\frac{1}{2}\theta_U.$$

In the low state, informed traders sell rather than buy, so

$$\mathbb{P}(B, Lg \mid v_L) = (1-\mu)\frac{1}{2}\theta_U.$$

Thus

$$\mathbb{P}(v_H \mid B, Lg) = \frac{p\{\mu\theta_I + (1-\mu)\theta_U/2\}}{p\{\mu\theta_I + (1-\mu)\theta_U/2\} + (1-p)(1-\mu)\theta_U/2}.$$

For a small buy,

$$\mathbb{P}(B, Sm \mid v_H) = \mu(1-\theta_I) + (1-\mu)\frac{1}{2}(1-\theta_U),$$

and

$$\mathbb{P}(B, Sm \mid v_L) = (1-\mu)\frac{1}{2}(1-\theta_U).$$

Therefore

$$\mathbb{P}(v_H \mid B, Sm) = \frac{p\{\mu(1-\theta_I) + (1-\mu)(1-\theta_U)/2\}}{p\{\mu(1-\theta_I) + (1-\mu)(1-\theta_U)/2\} + (1-p)(1-\mu)(1-\theta_U)/2}.$$

Because $\theta_I > \theta_U$, large orders are relatively more likely to come from informed traders. Hence a large buy pushes the posterior toward the high state more than a small buy, and the large-buy ask exceeds the small-buy ask.

An empirical reduced-form test is

$$\Delta m_t = \alpha + \lambda q_t + \gamma q_t \mathbf{1}_{\{V_t > \bar{V}\}} + \varepsilon_t.$$

If $\gamma > 0$, large signed trades have extra permanent price impact beyond the baseline trade-sign effect. □

Takeaway. *Trade size is not just volume. In adverse-selection models it is a mark on the order-flow process and can carry information about the likelihood that the trader is informed.*

References for this problem.

- Easley and O'Hara (1987), for price, trade size, and information.
- O'Hara (1995), for adverse-selection models with trade-size information.
- Glosten and Harris (1988), for empirical size-based spread decomposition.

2.2 Time, Silence, and PIN

Problem 2.4 (Advanced: No-trade intervals as information)**.** Suppose a day is an information-event day with prior probability α. On no-event days, uninformed trades arrive as a Poisson process with intensity ϵ. On event days, uninformed trades arrive with intensity ϵ and informed trades arrive with intensity μ. Let N_τ be the number of trades observed over an interval of length τ.

(a) Compute $\mathbb{P}(N_\tau = 0 \mid \text{event})$.

(b) Compute $\mathbb{P}(N_\tau = 0 \mid \text{no event})$.

(c) Derive $\mathbb{P}(\text{event} \mid N_\tau = 0)$.

(d) Explain why a no-trade interval can narrow spreads.

Solution.

On an event day, total arrival intensity is $\epsilon + \mu$, so

$$\mathbb{P}(N_\tau = 0 \mid \text{event}) = \exp\{-(\epsilon + \mu)\tau\}.$$

On a no-event day, total arrival intensity is ϵ, so

$$\mathbb{P}(N_\tau = 0 \mid \text{no event}) = \exp\{-\epsilon\tau\}.$$

By Bayes' rule,

$$\mathbb{P}(\text{event} \mid N_\tau = 0) = \frac{\alpha e^{-(\epsilon+\mu)\tau}}{\alpha e^{-(\epsilon+\mu)\tau} + (1-\alpha)e^{-\epsilon\tau}}.$$

Dividing numerator and denominator by $e^{-\epsilon\tau}$ gives

$$\mathbb{P}(\text{event} \mid N_\tau = 0) = \frac{\alpha e^{-\mu\tau}}{\alpha e^{-\mu\tau} + 1 - \alpha}.$$

This posterior decreases with τ. A long silence is evidence against an information event because informed traders would have been more likely to arrive if an event had occurred. If the probability of informed trading falls, the adverse-selection component of the spread can narrow. □

Takeaway. *In sequential-trade models, time is informative. A period with no trades is not missing data; it is evidence about the latent information state.*

References for this problem.

- Easley and O'Hara (1992), for time and no-trade intervals as information.
- O'Hara (1995), for sequential-trade models with time as an information variable.
- Engle and Russell (1998), for duration modeling of irregular transaction arrivals.

Problem 2.5 (Advanced: Deriving PIN from a mixture likelihood)**.** For each day d, observe the number of buyer-initiated and seller-initiated trades (B_d, S_d). Suppose there are three latent day types:

$$\text{no news}, \qquad \text{good news}, \qquad \text{bad news}.$$

An information event occurs with probability α. Conditional on an event, the news is bad with probability δ and good with probability $1 - \delta$. Uninformed buy and sell arrivals have intensities ϵ_b and ϵ_s. Informed arrivals have intensity μ and buy on good-news days, sell on bad-news days.

(a) Write the likelihood contribution for a no-news day.

(b) Write the likelihood contribution for a good-news day.

(c) Write the likelihood contribution for a bad-news day.

(d) Combine these into the daily mixture likelihood.

(e) Derive the probability of informed trading, PIN.

Solution.

Let

$$P(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!}$$

denote the Poisson probability. On a no-news day,

$$L_0(B_d, S_d) = P(B_d; \epsilon_b) P(S_d; \epsilon_s).$$

On a good-news day, informed traders buy, so

$$L_G(B_d, S_d) = P(B_d; \epsilon_b + \mu)P(S_d; \epsilon_s).$$

On a bad-news day, informed traders sell, so

$$L_B(B_d, S_d) = P(B_d; \epsilon_b)P(S_d; \epsilon_s + \mu).$$

The daily likelihood contribution is the mixture

$$L_d = (1 - \alpha)L_0(B_d, S_d) + \alpha(1 - \delta)L_G(B_d, S_d) + \alpha\delta L_B(B_d, S_d).$$

Across independent days, the likelihood is $\prod_d L_d$.

The expected informed-trade arrival intensity is $\alpha\mu$. The expected total trade-arrival intensity is

$$\alpha\mu + \epsilon_b + \epsilon_s.$$

Thus

$$PIN = \frac{\alpha\mu}{\alpha\mu + \epsilon_b + \epsilon_s}.$$

This object is a model-implied fraction of expected trades that are informed. □

Takeaway. *PIN is not an observed fraction. It is a latent-mixture estimate: a model turns daily buy and sell counts into an inferred information-event intensity.*

References for this problem.

- Easley, Kiefer, O'Hara, and Paperman (1996), for the EKOP/PIN mixture likelihood.
- Easley, Kiefer, and O'Hara (1997), for empirical sequential-trade/PIN estimation.
- Hasbrouck (2007), for the empirical exposition of PIN models.
- Lin and Ke (2011), for numerical instability in PIN estimation.

Problem 2.6 (Research: Numerically stable PIN likelihood)**.** The raw EKOP likelihood contains Poisson probabilities with large count factorials and exponential terms. For day d, define

$$L_d = (1 - \alpha)L_0 + \alpha(1 - \delta)L_G + \alpha\delta L_B,$$

where

$$L_0 = P(B_d; \epsilon_b)P(S_d; \epsilon_s),$$

$$L_G = P(B_d; \epsilon_b + \mu)P(S_d; \epsilon_s), \qquad L_B = P(B_d; \epsilon_b)P(S_d; \epsilon_s + \mu),$$

and $P(k;\lambda) = e^{-\lambda}\lambda^k/k!$.

(a) Explain why direct numerical evaluation of L_d can underflow or overflow when counts are large.

(b) Factor out the common no-news Poisson term $P(B_d;\epsilon_b)P(S_d;\epsilon_s)$.

(c) Derive a stable log-likelihood contribution using the log-sum-exp operator.

(d) Explain why this numerical reformulation does not change the model.

Solution.

Direct evaluation is unstable because the terms contain $B_d!$, $S_d!$, large powers such as $(\epsilon_b+\mu)^{B_d}$, and exponentials such as $e^{-(\epsilon_b+\mu)}$. With high-frequency counts, these components can be individually enormous or tiny even when their product is economically ordinary.

Let

$$C_d = P(B_d;\epsilon_b)P(S_d;\epsilon_s).$$

For the good-news term,

$$\frac{L_G}{C_d} = \frac{P(B_d;\epsilon_b+\mu)}{P(B_d;\epsilon_b)} = e^{-\mu}\left(1+\frac{\mu}{\epsilon_b}\right)^{B_d}.$$

For the bad-news term,

$$\frac{L_B}{C_d} = \frac{P(S_d;\epsilon_s+\mu)}{P(S_d;\epsilon_s)} = e^{-\mu}\left(1+\frac{\mu}{\epsilon_s}\right)^{S_d}.$$

Thus

$$L_d = C_d\left[(1-\alpha)+\alpha(1-\delta)e^{-\mu}\left(1+\frac{\mu}{\epsilon_b}\right)^{B_d}+\alpha\delta e^{-\mu}\left(1+\frac{\mu}{\epsilon_s}\right)^{S_d}\right].$$

The log of the common term is

$$\log C_d = -\epsilon_b-\epsilon_s+B_d\log\epsilon_b+S_d\log\epsilon_s-\log\Gamma(B_d+1)-\log\Gamma(S_d+1).$$

Define

$$a_0 = \log(1-\alpha),$$

$$a_G = \log\alpha+\log(1-\delta)-\mu+B_d\log\left(1+\frac{\mu}{\epsilon_b}\right),$$

and

$$a_B = \log\alpha+\log\delta-\mu+S_d\log\left(1+\frac{\mu}{\epsilon_s}\right).$$

Then the stable contribution is

$$\log L_d = \log C_d + \mathrm{LSE}(a_0, a_G, a_B),$$

where

$$\mathrm{LSE}(a_0, a_G, a_B) = m + \log\{e^{a_0 - m} + e^{a_G - m} + e^{a_B - m}\}, \qquad m = \max\{a_0, a_G, a_B\}.$$

Subtracting m keeps the exponentials between zero and one.

This is an algebraic reformulation, not a different estimator. The likelihood surface is unchanged; only the numerical representation is changed. The econometric lesson is that structural likelihoods must be written in a form that computers can evaluate reliably. □

Takeaway. *PIN estimation is not only about economic identification. It is also a numerical econometrics problem: the same likelihood can be unusable in raw Poisson form and stable after factorization and log-sum-exp evaluation.*

References for this problem.

- Lin and Ke (2011), for computing bias and numerical instability in PIN estimation.
- Easley, Hvidkjaer, and O'Hara (2002), for empirical PIN implementation and likelihood factorization.
- Easley, Kiefer, O'Hara, and Paperman (1996), for the original EKOP/PIN likelihood.
- Andersen and Bondarenko (2014), for caution on PIN/VPIN-style measurement.

2.3 Structural Sequential-Trade Econometrics

The next problems move from Bayesian quote setting to econometric implementation. They are inspired by the sequential-trade econometrics literature: latent information regimes, Poisson and negative-binomial arrival models, EM estimation, intraday seasonality, and specification testing.

Problem 2.7 (Advanced: Posterior regime probabilities in the EKOP count model)**.** For day d, observe buyer-initiated and seller-initiated trade counts (B_d, S_d). Let the latent regime be

$$Z_d \in \{0, G, B\},$$

where 0 is no news, G is good news, and B is bad news. The prior regime probabilities are

$$\pi_0 = 1 - \alpha, \qquad \pi_G = \alpha(1 - \delta), \qquad \pi_B = \alpha\delta.$$

Conditional on Z_d, buy and sell counts are independent Poisson random variables with intensities

$$(\lambda_0^B, \lambda_0^S) = (\epsilon_b, \epsilon_s), \quad (\lambda_G^B, \lambda_G^S) = (\epsilon_b + \mu, \epsilon_s), \quad (\lambda_B^B, \lambda_B^S) = (\epsilon_b, \epsilon_s + \mu).$$

(a) Derive the posterior probability $\mathbb{P}(Z_d = j \mid B_d, S_d)$.

(b) Show how the posterior classifies a day as good news or bad news.

(c) Explain why the likelihood is a finite mixture.

(d) Give an economic interpretation of posterior regime classification.

Solution.

Let

$$P(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

The regime-specific likelihood is

$$L_j(B_d, S_d) = P(B_d; \lambda_j^B) P(S_d; \lambda_j^S).$$

Bayes' rule gives the posterior responsibility

$$\tau_{dj} = \mathbb{P}(Z_d = j \mid B_d, S_d) = \frac{\pi_j L_j(B_d, S_d)}{\sum_{\ell \in \{0,G,B\}} \pi_\ell L_\ell(B_d, S_d)}.$$

A day is naturally classified as regime

$$\widehat{Z}_d = \arg \max_{j \in \{0,G,B\}} \tau_{dj}.$$

Large buy counts relative to sell counts raise L_G because the good-news regime has buy intensity $\epsilon_b + \mu$. Large sell counts raise L_B. Balanced counts with ordinary total activity raise L_0.

The likelihood is a finite mixture because the observed count vector is drawn from one of several latent regimes, but the econometrician does not observe which one. The posterior probabilities translate daily order-flow imbalance into probabilistic statements about latent information events.
□

Takeaway. *The EKOP likelihood is not just a formula for PIN. It is also a daily classifier of latent information states.*

References for this problem.

- Easley, Kiefer, O'Hara, and Paperman (1996), for posterior information regime probabilities from buy and sell counts.
- Easley, Kiefer, and O'Hara (1997), for empirical sequential-trade/PIN estimation.
- Hasbrouck (2007), for a first reading on PIN count models.
- Kokot (2004), for sequential-trade econometric extensions.

Problem 2.8 (Advanced: Identification from regime-specific arrival rates)**.** Suppose a reduced-form three-regime Poisson mixture estimates the buy and sell arrival-rate pairs

$$(\lambda_1^B, \lambda_1^S), \quad (\lambda_2^B, \lambda_2^S), \quad (\lambda_3^B, \lambda_3^S),$$

without imposing the EKOP restrictions.

(a) State conditions under which the regimes can be labeled as no news, good news, and bad news.

(b) Recover $\epsilon_b, \epsilon_s, \mu$ when the EKOP restrictions hold.

(c) Explain why a fourth regime is easier to interpret in a reduced-form mixture than in the structural EKOP model.

(d) Describe one empirical pattern that would reject the strict EKOP restrictions.

Solution.

The no-news regime should have approximately balanced ordinary flow:

$$\lambda_j^B \approx \epsilon_b, \qquad \lambda_j^S \approx \epsilon_s.$$

The good-news regime has an excess buy intensity, so

$$\lambda_j^B - \lambda_j^S$$

should be large and positive after accounting for baseline asymmetry. The bad-news regime has a large negative imbalance.

If regime 0 is no news, regime G good news, and regime B bad news, then

$$\epsilon_b = \lambda_0^B, \qquad \epsilon_s = \lambda_0^S,$$

and the informed intensity can be recovered as

$$\mu = \lambda_G^B - \lambda_0^B = \lambda_B^S - \lambda_0^S.$$

The equality of these two estimates is itself a restriction.

A fourth reduced-form regime simply adds another pair $(\lambda_4^B, \lambda_4^S)$ and lets the data reveal whether it corresponds to extreme good news, extreme bad news, a high-liquidity balanced state, or some other state. In the structural EKOP model, a fourth regime has no immediate theoretical label without adding new economic assumptions.

The strict model is rejected if, for example,

$$\lambda_G^B - \lambda_0^B \neq \lambda_B^S - \lambda_0^S,$$

or if the no-news regime itself shows strong directional imbalance. □

Takeaway. *Reduced-form mixtures sacrifice some structural interpretation but gain a cleaner way to discover how many trading regimes the data actually support.*

References for this problem.

- Easley, Kiefer, O'Hara, and Paperman (1996), for identifying informed and uninformed arrival rates in EKOP.
- Easley, Kiefer, and O'Hara (1997), for empirical arrival-rate estimation.
- Hasbrouck (2007), for interpretation of PIN parameters.
- Kokot (2004), for structural identification in sequential-trade models.

Problem 2.9 (Advanced: EM estimation of a three-regime sequential-trade model)**.** Let $Y_d = (B_d, S_d)$ and let $Z_d \in \{1, 2, 3\}$ be an unobserved regime. Conditional on regime j,

$$B_d \sim \text{Poisson}(\lambda_j^B), \qquad S_d \sim \text{Poisson}(\lambda_j^S),$$

independently. The regime probability is π_j.

(a) Write the observed-data likelihood.

(b) Derive the E-step posterior responsibility τ_{dj}.

(c) Derive the M-step updates for π_j, λ_j^B, and λ_j^S.

(d) Explain why the EM algorithm is numerically natural for this problem.

Solution.

The observed-data likelihood contribution is

$$L_d = \sum_{j=1}^{3} \pi_j P(B_d; \lambda_j^B) P(S_d; \lambda_j^S).$$

Thus the sample likelihood is $\prod_d L_d$.

Given current parameters, the E-step computes

$$\tau_{dj} = \frac{\pi_j P(B_d; \lambda_j^B) P(S_d; \lambda_j^S)}{\sum_{\ell=1}^{3} \pi_\ell P(B_d; \lambda_\ell^B) P(S_d; \lambda_\ell^S)}.$$

The M-step maximizes the expected complete-data log likelihood. The updates are

$$\pi_j^{new} = \frac{1}{D}\sum_{d=1}^{D} \tau_{dj},$$

$$(\lambda_j^B)^{new} = \frac{\sum_{d=1}^{D} \tau_{dj} B_d}{\sum_{d=1}^{D} \tau_{dj}}, \qquad (\lambda_j^S)^{new} = \frac{\sum_{d=1}^{D} \tau_{dj} S_d}{\sum_{d=1}^{D} \tau_{dj}}.$$

So each arrival rate is a posterior-probability-weighted sample mean.

EM is natural because the difficult part of the likelihood is the unobserved regime. Conditional on posterior regime weights, the Poisson parameters have simple closed-form updates. □

Takeaway. *EM turns latent-regime estimation into repeated soft classification followed by weighted count estimation.*

References for this problem.

- Dempster, Laird, and Rubin (1977), for the EM algorithm.
- Easley, Kiefer, O'Hara, and Paperman (1996), for the latent-regime EKOP count mixture.
- Hamilton (1994), for likelihood estimation with latent states.
- Lin and Ke (2011), for numerical issues when implementing PIN likelihoods.

Problem 2.10 (Advanced: Poisson regression mixture with intraday seasonality)**.** Partition each trading day into intervals $t = 1, \ldots, T$. Let Y_{tk} be the count of event type $k \in \{B, S\}$ in interval t. Conditional on latent regime $Z_t = j$,

$$Y_{tk} \sim \text{Poisson}(\Delta_t \lambda_{tjk}), \qquad \log \lambda_{tjk} = x_t' \beta_{jk},$$

where x_t includes time-of-day terms.

(a) Write the mixture likelihood contribution for interval t.

(b) Derive the E-step responsibility.

(c) Show why the M-step for β_{jk} is a weighted Poisson regression.

(d) Explain why intraday seasonality must be controlled before interpreting regimes as information states.

Solution.

The regime-specific density is

$$f_j(Y_t \mid x_t) = \prod_{k\in\{B,S\}} P(Y_{tk}; \Delta_t \exp(x_t'\beta_{jk})).$$

The interval likelihood is

$$L_t = \sum_j \pi_j f_j(Y_t \mid x_t).$$

The E-step gives

$$\tau_{tj} = \frac{\pi_j f_j(Y_t \mid x_t)}{\sum_\ell \pi_\ell f_\ell(Y_t \mid x_t)}.$$

The expected complete-data log likelihood separates across j and k:

$$\sum_t \tau_{tj}\{Y_{tk}x_t'\beta_{jk} - \Delta_t \exp(x_t'\beta_{jk})\} + \text{constant}.$$

This is exactly the Poisson regression objective with observation weights τ_{tj} and exposure Δ_t.

Without seasonality controls, the model may interpret the market open and close as information regimes simply because trading intensity is naturally high there. Time-of-day controls separate deterministic liquidity cycles from latent information states. □

Takeaway. *In high-frequency data, regime detection without a time-of-day adjustment is often just seasonality detection in disguise.*

References for this problem.

- Easley, Kiefer, O'Hara, and Paperman (1996), for the Poisson count mixture benchmark.
- Kokot (2004), for sequential-trade count models with time-varying intensities.
- Cameron and Trivedi (2013), for Poisson regression foundations.
- Engle and Russell (1998), for duration-based modeling of irregular transaction timing.

Problem 2.11 (Advanced: Fourier seasonality in arrival intensities)**.** Let $\tau_t \in [0, 1]$ denote normalized time of day. Consider the seasonal specification

$$s(\tau_t) = a_0 + a_1\tau_t + a_2\tau_t^2 + c_1 \cos(2\pi\tau_t) + d_1 \sin(2\pi\tau_t),$$

and

$$\log \lambda_{tjk} = s(\tau_t) + z_t'\gamma_{jk}.$$

(a) Explain why a Fourier-polynomial form is useful for intraday intensity.

(b) Derive the multiplicative effect of seasonality on λ_{tjk}.

(c) Show how one would test whether seasonality is needed.

(d) Explain the identification issue if each regime has its own unrestricted seasonality function.

Solution.

The intraday intensity curve is typically high near the open and close and lower in the middle of the day. A polynomial plus sine and cosine terms can capture this smooth U-shape while respecting the bounded support of the trading day.

Because the model is log-linear,

$$\lambda_{tjk} = \exp\{s(\tau_t)\}\exp\{z_t'\gamma_{jk}\}.$$

Seasonality therefore scales the baseline arrival intensity multiplicatively.

A likelihood-ratio test compares the unrestricted model to the restricted model

$$a_1 = a_2 = c_1 = d_1 = 0.$$

One can also use information criteria if the model is a mixture and regular likelihood theory is delicate.

If every regime has a completely different seasonal function, a high-intensity time of day can become confounded with a high-intensity regime. A conservative specification uses common seasonal controls and lets regimes explain residual differences in buy and sell arrival behavior. □

Takeaway. *Seasonality is not a nuisance detail. It determines whether a regime model is finding information events or simply finding the clock.*

References for this problem.

- Easley, Kiefer, O'Hara, and Paperman (1996), for the arrival-rate mixture model.
- Kokot (2004), for seasonal intensity extensions in sequential-trade econometrics.
- Cameron and Trivedi (2013), for Poisson regression with covariates.
- Hasbrouck (2007), for intraday empirical implementation issues.

Problem 2.12 (Advanced: Negative-binomial mixtures and trader heterogeneity)**.** Suppose the Poisson mixture underfits the variance of buy and sell counts. For event type k in regime j, replace the Poisson model with a negative-binomial model satisfying

$$\mathbb{E}[Y_{tk} \mid Z_t = j] = m_{tjk}, \qquad \mathrm{Var}(Y_{tk} \mid Z_t = j) = m_{tjk} + a_{jk}m_{tjk}^2.$$

(a) Explain why the Poisson model cannot capture overdispersion.

(b) Interpret the parameter a_{jk} economically.

(c) State how the E-step changes relative to the Poisson mixture.

(d) Explain how overdispersion can affect PIN estimates.

Solution.

The Poisson restriction is

$$\mathrm{Var}(Y_{tk} \mid Z_t = j) = \mathbb{E}[Y_{tk} \mid Z_t = j].$$

High-frequency trade counts often have variance much larger than the mean, due to clustering, heterogeneity in participants, common shocks, and unmodeled strategic behavior.

The parameter a_{jk} measures extra variance beyond Poisson variation. In a microstructure interpretation, it captures within-regime heterogeneity in trading intensities: traders in the same broad information state need not have identical arrival rates.

The E-step formula is the same Bayes formula,

$$\tau_{tj} = \frac{\pi_j f_j^{NB}(Y_t)}{\sum_\ell \pi_\ell f_\ell^{NB}(Y_t)},$$

but the regime density f_j^{NB} is negative-binomial rather than Poisson. The M-step no longer reduces to simple Poisson weighted means, but it is still a weighted count-regression problem.

If overdispersion is ignored, extreme count days may be falsely interpreted as information-event days. That can push $\widehat{\alpha}$ or $\widehat{\mu}$ upward and inflate estimated PIN. □

Takeaway. *Extra count variance is economically meaningful. It can be trader heterogeneity, not necessarily more informed trading.*

References for this problem.

- Easley, Kiefer, O'Hara, and Paperman (1996), for the Poisson mixture benchmark.
- Cameron and Trivedi (2013), for negative-binomial overdispersion.
- Kokot (2004), for overdispersion and sequential-trade applications.

2.4 Diagnostics, Regime Dynamics, and Data Construction

Problem 2.13 (Advanced: Markov-switching trade regimes)**.** Let the information regime $Z_t \in \{1, \dots, N\}$ follow a Markov chain with transition probabilities

$$p_{ij} = \mathbb{P}(Z_t = j \mid Z_{t-1} = i).$$

Conditional on $Z_t = j$, the event-count vector has density $f_j(Y_t)$.

(a) Derive the one-step predicted regime probability.

(b) Derive the filtered probability after observing Y_t.

(c) Show how the static mixture is nested in the Markov model.

(d) Explain why regime persistence is natural for information events.

Solution.

Let

$$\xi_{t-1|t-1}(i) = \mathbb{P}(Z_{t-1} = i \mid \mathcal{F}_{t-1}).$$

The predicted probability is

$$\xi_{t|t-1}(j) = \sum_i p_{ij}\xi_{t-1|t-1}(i).$$

After observing Y_t,

$$\xi_{t|t}(j) = \frac{f_j(Y_t)\xi_{t|t-1}(j)}{\sum_{\ell=1}^N f_\ell(Y_t)\xi_{t|t-1}(\ell)}.$$

The static mixture is nested by setting

$$p_{ij} = \pi_j \quad \text{for every } i.$$

Then the next regime is independent of the previous regime.

Regime persistence is natural because information events, trading interest, and liquidity states do not usually reset independently every minute. A news state can persist across adjacent intervals, creating autocorrelation in counts. □

Takeaway. *Static mixtures classify observations. Markov mixtures classify histories.*

References for this problem.

- Hamilton (1989), for Markov regime switching.
- Hamilton (1994), for the textbook treatment of Markov-switching likelihoods.
- Easley, Kiefer, O'Hara, and Paperman (1996), for the static latent information-event benchmark.

Problem 2.14 (Advanced: Testing the number of information regimes)**.** A researcher estimates finite mixtures with $N = 2, 3, 4$ regimes for signed trade-count vectors.

(a) Explain why the number of regimes is difficult to test in a structural EKOP specification.

(b) Give an information-criterion approach to choosing N.

(c) Explain label switching and why it matters for interpretation.

(d) Describe one economic reason a fourth regime may appear.

Solution.

In the structural model, each regime has an economic meaning. Moving from three to four regimes requires deciding what the fourth regime means before estimation. Different restrictions lead to different tests, so the testing problem is not neutral.

In a reduced-form mixture, estimate each N and compute, for example,

$$BIC_N = -2\ell_N + k_N \log T,$$

where ℓ_N is the maximized log likelihood and k_N the number of parameters. The preferred model has the lowest BIC, subject to economic interpretability.

Label switching means the likelihood is unchanged if regime names are permuted. The econometric output must be relabeled after estimation, usually by ordering regimes according to buy-sell imbalance or total intensity.

A fourth regime might represent high-volume balanced trading, extreme news, liquidity demand unrelated to information, or an opening/closing auction effect that was not fully absorbed by seasonality controls. □

Takeaway. *Model selection in sequential-trade econometrics is both statistical and economic: a better likelihood is useful only if the added regime means something.*

References for this problem.

- Hansen (1982), for moment-based specification testing.
- Dempster, Laird, and Rubin (1977), for mixture likelihood estimation.
- Hamilton (1994), for model comparison with latent states.
- Newey and West (1987), for HAC covariance estimation under dependence.

Problem 2.15 (Research: Trade-direction misclassification and PIN attenuation)**.** Suppose true buys and sells are (B_d, S_d), but the econometrician observes classified counts $(\widetilde{B}_d, \widetilde{S}_d)$. Each buy is incorrectly classified as a sell with probability η, and each sell is incorrectly classified as a buy with the same probability η.

(a) Compute $\mathbb{E}[\widetilde{B}_d \mid B_d, S_d]$ and $\mathbb{E}[\widetilde{S}_d \mid B_d, S_d]$.

(b) Show how misclassification shrinks observed order imbalance.

(c) Explain the likely direction of bias in $\widehat{\mu}$.

(d) Propose one correction.

Solution.

The observed buy count is true buys that remain buys plus true sells classified as buys:

$$\mathbb{E}[\widetilde{B}_d \mid B_d, S_d] = (1-\eta)B_d + \eta S_d.$$

Similarly,

$$\mathbb{E}[\widetilde{S}_d \mid B_d, S_d] = \eta B_d + (1-\eta)S_d.$$

Therefore

$$\mathbb{E}[\widetilde{B}_d - \widetilde{S}_d \mid B_d, S_d] = (1-2\eta)(B_d - S_d).$$

When $0 < \eta < 1/2$, observed imbalance is attenuated toward zero.

Since informed trading in the EKOP model appears as excess buys on good-news days and excess sells on bad-news days, attenuation of imbalance tends to reduce the estimated informed arrival intensity μ. A correction is to include the classification matrix in the likelihood, so the likelihood of observed counts integrates over possible true buy and sell counts. A simpler robustness check is to re-estimate the model under plausible fixed values of η. □

Takeaway. *PIN estimates depend on market microstructure data construction. Trade signing is not a clerical step; it is part of the econometric model.*

References for this problem.

- Lee and Ready (1991), for trade-direction inference from intraday data.
- Easley, Kiefer, O'Hara, and Paperman (1996), for PIN estimation using signed buy and sell counts.
- Hasbrouck (2007), for empirical trade signing and PIN implementation.
- Lin and Ke (2011), for PIN implementation sensitivity.

Problem 2.16 (Advanced: Conditional moment test for regime models)**.** Let $\widehat{m}_t = \mathbb{E}[Y_t \mid \mathcal{F}_{t-1}; \widehat{\theta}]$ be the fitted conditional mean from a sequential-trade count model. Define the residual

$$u_t = Y_t - \widehat{m}_t.$$

(a) State the moment restriction implied by correct conditional-mean specification.

(b) Construct a test using instruments w_t known at $t-1$.

(c) Explain how this test can reveal omitted seasonality or regime dependence.

(d) Give one limitation of the test.

Solution.

Correct conditional-mean specification implies

$$\mathbb{E}[u_t \mid \mathcal{F}_{t-1}] = 0.$$

Therefore, for any instrument w_t measurable with respect to $\mathcal{F}_{t-1}$,

$$\mathbb{E}[w_t u_t] = 0.$$

The sample moment is

$$\bar{g}_T = \frac{1}{T}\sum_{t=1}^{T} w_t u_t.$$

A quadratic-form test uses

$$J = T\bar{g}_T' \widehat{W}^{-1} \bar{g}_T,$$

where $\widehat{W}$ estimates the long-run covariance of $w_t u_t$.

If the model omits seasonality, instruments such as time-of-day dummies will be correlated with residuals. If it omits regime persistence, lagged counts or lagged residuals may forecast u_t.

The limitation is that rejection does not identify a unique fix. A failed moment test may reflect seasonality, overdispersion, serial dependence, misclassification, or an incorrect structural model. □

Takeaway. *Specification testing should ask whether the model has exhausted predictable structure in order flow. If residual counts are still forecastable, the sequential-trade model is incomplete.*

References for this problem.

- Hansen (1982), for generalized method of moments.
- Newey and West (1987), for HAC covariance estimation.
- Hamilton (1994), for specification testing in time-series settings.
- Easley, Kiefer, O'Hara, and Paperman (1996), for the regime model whose implications are being tested.

Problem 2.17 (Research: Aggregation interval and the no-trade signal)**.** A continuous-time sequential-trade process is observed in intervals of length Δ. In each interval the econometrician records counts of buys and sells, but not the exact arrival times.

(a) Explain what information is lost when exact event times are aggregated into counts.

(b) Derive the probability of no trade in an interval with total intensity λ.

(c) Discuss the bias-variance tradeoff in choosing Δ.

(d) Explain why very long intervals can weaken tests of no-trade information.

Solution.

Aggregation loses the order and spacing of trades inside the interval. Two intervals with the same buy and sell counts may have very different event-time patterns, and the timing itself can carry information about latent states.

For a Poisson process with total intensity λ,

$$\mathbb{P}(N_\Delta = 0) = e^{-\lambda\Delta}.$$

Short intervals preserve no-trade information and timing variation, but they produce many zeros and noisy estimates. Long intervals stabilize counts, but they smooth away the distinction between an early burst followed by silence and a steady arrival process.

If Δ is too long, no-trade events become rare except in very illiquid assets. The econometrician then loses the silence signal that sequential-trade models treat as evidence about the absence of informed trading. □

Takeaway. *Sampling frequency is a modeling choice. Sequential-trade econometrics is about events in time, not merely daily totals.*

References for this problem.

- Easley and O'Hara (1992), for no-trade intervals as information.
- Easley, Kiefer, O'Hara, and Paperman (1996), for fixed-interval buy/sell count aggregation.
- Hasbrouck (2007), for PIN implementation and transaction-time aggregation.
- Engle and Russell (1998), for duration modeling as an alternative to fixed-interval aggregation.

2.5 Flow Toxicity, VPIN, and Generalized EKOP Models

The final block in this chapter connects structural sequential-trade models to the later flow-toxicity literature. The goal is not to treat VPIN as a magic crash predictor, but to understand when volume imbalance is a useful empirical summary and when it is only a noisy reduced-form proxy for latent adverse selection.

Problem 2.18 (Advanced: VPIN as volume-bucket order imbalance)**.** Trades are grouped into volume buckets of equal size V. In bucket τ, let B_τ and S_τ denote estimated buy and sell volume, with

$$B_\tau + S_\tau = V.$$

Define the volume imbalance

$$I_\tau = |B_\tau - S_\tau|.$$

For a rolling window of n buckets, define

$$VPIN_t = \frac{1}{nV} \sum_{\tau=t-n+1}^{t} I_\tau.$$

(a) Show that $0 \le VPIN_t \le 1$.

(b) Express I_τ in terms of the buy-volume share $x_\tau = B_\tau/V$.

(c) Explain why VPIN rises when volume becomes one-sided.

(d) Give two reasons why VPIN is only a proxy for toxicity, not a direct estimate of informed trading.

Solution.

Because $B_\tau, S_\tau \ge 0$ and $B_\tau + S_\tau = V$,

$$0 \le |B_\tau - S_\tau| \le V.$$

Therefore

$$0 \le \frac{1}{nV} \sum_{\tau=t-n+1}^{t} I_\tau \le 1.$$

Let $x_\tau = B_\tau/V$. Then $S_\tau/V = 1 - x_\tau$, so

$$\frac{I_\tau}{V} = |x_\tau - (1 - x_\tau)| = |2x_\tau - 1|.$$

Thus

$$VPIN_t = \frac{1}{n} \sum_{\tau=t-n+1}^{t} |2x_\tau - 1|.$$

VPIN rises when buckets are dominated by buys or dominated by sells. It is low when volume is balanced. This is why the statistic is often interpreted as a measure of order-flow toxicity: a market maker facing one-sided volume may be trading against better-informed counterparties or against urgent directional liquidity demand.

But VPIN is not a structural probability of informed trading. First, buy and sell volume are often estimated rather than directly observed, so classification error enters mechanically. Second, one-sided flow can arise from inventory shocks, index rebalancing, execution algorithms, stop-loss cascades, or forced liquidation, not only private information. □

Takeaway. *VPIN is a useful reduced-form imbalance statistic. Its economic interpretation requires a model of why order flow became one-sided.*

References for this problem.

- Easley, Lopez de Prado, and O'Hara (2012), for VPIN and flow toxicity.
- Hasbrouck (2007), for PIN background before the volume-bucket extension.
- Andersen and Bondarenko (2014), for the critique of VPIN as a flash-crash warning statistic.

Problem 2.19 (Research: Flow toxicity and flash-crash fragility)**.** Consider a market maker who sets depth D_t inversely to perceived flow toxicity:

$$D_t = D_0 \exp\{-\kappa\, VPIN_t\}, \qquad \kappa > 0.$$

Suppose signed net volume over a short interval is $Q_t = B_t - S_t$ and price impact is locally

$$\Delta p_t = \frac{Q_t}{D_t}.$$

(a) Derive $\partial D_t / \partial VPIN_t$.

(b) Derive the price impact of a fixed sell imbalance $Q_t < 0$ as VPIN rises.

(c) Explain how this mechanism can amplify a flash-crash episode.

(d) Explain why high VPIN is not sufficient to prove that informed traders caused the crash.

Solution.

Differentiating depth with respect to VPIN gives

$$\frac{\partial D_t}{\partial VPIN_t} = -\kappa D_0 \exp\{-\kappa VPIN_t\} = -\kappa D_t < 0.$$

Thus perceived toxicity reduces displayed or effective depth.

For a fixed imbalance Q_t,

$$\Delta p_t = \frac{Q_t}{D_0 \exp\{-\kappa VPIN_t\}} = \frac{Q_t}{D_0} \exp\{\kappa VPIN_t\}.$$

If $Q_t < 0$, the price change becomes more negative in magnitude as VPIN rises.

This creates an amplification channel. One-sided flow raises VPIN; higher VPIN reduces depth; lower depth makes the next imbalance move prices more; larger price moves may trigger further order-flow imbalance through stop-losses, margin constraints, or liquidity withdrawal.

However, high VPIN alone does not prove informed trading caused the crash. It may reflect classification errors, endogenous liquidity withdrawal, mechanical execution pressure, correlated liquidity demand, or feedback trading. A causal claim requires additional evidence about information arrival, trader identity, order-book depth, cancellations, and alternative liquidity-demand shocks. □

Takeaway. *Flow toxicity can be a fragility signal even when it is not a pure information signal. In crashes, adverse selection and liquidity spirals are empirically hard to separate.*

References for this problem.

- Easley, Lopez de Prado, and O'Hara (2012), for flow toxicity in high-frequency markets.
- O'Hara (1995), for adverse-selection foundations behind liquidity withdrawal.
- Andersen and Bondarenko (2014), for critical evidence on VPIN and the Flash Crash.

Problem 2.20 (Research: A generalized EKOP model with volume buckets)**.** Extend the EKOP model from daily buy and sell counts to volume buckets. In bucket τ, observe buy and sell volume (B_τ, S_τ) with total volume $V_\tau = B_\tau + S_\tau$. Let the latent regime

$$Z_\tau \in \{0, G, B, H\}$$

represent no news, good news, bad news, and high-liquidity balanced trading. Conditional on $Z_\tau = j$, assume

$$B_\tau \sim \text{Poisson}(V_\tau \theta_j^B), \qquad S_\tau \sim \text{Poisson}(V_\tau \theta_j^S),$$

independently, where V_τ acts as exposure.

(a) Write the mixture likelihood contribution for bucket τ.

(b) Derive the posterior probability $\mathbb{P}(Z_\tau = j \mid B_\tau, S_\tau)$.

(c) State restrictions that recover an EKOP-like three-regime model.

(d) Explain how VPIN can be interpreted as a reduced-form statistic from this generalized model.

(e) Explain what the fourth regime H adds.

Solution.

Let $\pi_j = \mathbb{P}(Z_\tau = j)$ and let

$$P(k;\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

The regime-specific likelihood is

$$L_{\tau j} = P(B_\tau; V_\tau \theta_j^B) P(S_\tau; V_\tau \theta_j^S).$$

The mixture likelihood contribution is

$$L_\tau = \sum_{j\in\{0,G,B,H\}} \pi_j L_{\tau j}.$$

The posterior regime probability is

$$\tau_{\tau j} = \mathbb{P}(Z_\tau = j \mid B_\tau, S_\tau) = \frac{\pi_j L_{\tau j}}{\sum_{\ell\in\{0,G,B,H\}} \pi_\ell L_{\tau\ell}}.$$

An EKOP-like three-regime structure is recovered by dropping H and imposing

$$\theta_0^B = \epsilon_b, \qquad \theta_0^S = \epsilon_s,$$

$$\theta_G^B = \epsilon_b + \mu, \qquad \theta_G^S = \epsilon_s,$$

$$\theta_B^B = \epsilon_b, \qquad \theta_B^S = \epsilon_s + \mu.$$

The generalized model lets the intensities scale with bucket exposure V_τ, so buckets of different size can still be compared through arrival rates per unit volume.

VPIN summarizes the magnitude of realized imbalance,

$$\frac{|B_\tau - S_\tau|}{B_\tau + S_\tau},$$

while the generalized EKOP model asks which latent regime most likely produced that imbalance. A high VPIN bucket should raise posterior mass on G or B if one-sided flow is best explained by information, but it may raise posterior mass on H or another liquidity regime if high volume is balanced or non-informational.

The fourth regime H is useful because many high-volume periods are not news events. They may be index rebalancing, open/close effects, algorithmic execution, or broad liquidity demand. Without H, the model may force high-intensity but non-informational buckets into good-news or bad-news states. □

Takeaway. *VPIN and EKOP answer different questions. VPIN measures imbalance; generalized EKOP tries to explain imbalance through latent economic regimes.*

References for this problem.

- Easley, Kiefer, O'Hara, and Paperman (1996), for the EKOP regime mixture model.
- Easley, Lopez de Prado, and O'Hara (2012), for VPIN volume buckets.
- Hasbrouck (2007), for PIN and adverse-selection econometrics.
- Andersen and Bondarenko (2014), for critique of volume-bucket toxicity measures.

2.6 Modern Diagnostics: Covariates, Hawkes Toxicity, and Crash Warnings

This short closing block makes the adverse-selection chapter more current. It shows how structural sequential-trade models can be enriched with covariates, self-exciting order flow, and explicit out-of-sample diagnostic discipline.

Problem 2.21 (Research: Generalized EKOP with covariates and time-varying intensities)**.** For each interval t, observe buy and sell counts (B_t, S_t) and covariates X_t such as spread, depth imbalance, realized volatility, and time-of-day dummies. Let the latent regime be $Z_t \in \{0, G, B\}$: no news, good news, and bad news. Conditional on $Z_t = j$ and X_t, assume

$$B_t \sim \text{Poisson}(\lambda_j^B(X_t)), \qquad S_t \sim \text{Poisson}(\lambda_j^S(X_t)),$$

independently, with

$$\lambda_j^B(X_t) = \exp(a_j^B + X_t'\beta_j^B), \qquad \lambda_j^S(X_t) = \exp(a_j^S + X_t'\beta_j^S).$$

(a) Write the conditional mixture likelihood contribution.

(b) Derive the posterior probability $\mathbb{P}(Z_t = j \mid B_t, S_t, X_t)$.

(c) State coefficient restrictions that recover the classic EKOP interpretation of good and bad news.

(d) Explain how time-varying intensities help separate information from ordinary intraday activity.

Solution.

Let $\pi_j = \mathbb{P}(Z_t = j)$ and define

$$P(k;\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

The regime-specific likelihood is

$$L_{tj} = P(B_t;\lambda_j^B(X_t))P(S_t;\lambda_j^S(X_t)).$$

The conditional mixture likelihood contribution is

$$L_t = \sum_{j\in\{0,G,B\}} \pi_j L_{tj}.$$

The posterior regime probability is

$$\tau_{tj} = \mathbb{P}(Z_t = j \mid B_t, S_t, X_t) = \frac{\pi_j L_{tj}}{\sum_{\ell\in\{0,G,B\}} \pi_\ell L_{t\ell}}.$$

A classic EKOP interpretation imposes directional restrictions. For good news, buy intensity should exceed the no-news buy intensity while sell intensity should be close to its no-news level:

$$\lambda_G^B(X_t) > \lambda_0^B(X_t), \qquad \lambda_G^S(X_t) \approx \lambda_0^S(X_t).$$

For bad news,

$$\lambda_B^S(X_t) > \lambda_0^S(X_t), \qquad \lambda_B^B(X_t) \approx \lambda_0^B(X_t).$$

In a log-linear model these restrictions can be imposed or checked through the intercepts and covariate slopes.

Time-varying intensities are important because high trading activity is not always information. Opens, closes, macro-announcement windows, index rebalancing, and volatility spikes can raise both buy and sell counts. By conditioning on X_t, the model asks whether order flow is one-sided relative to what ordinary market conditions predict. □

Takeaway. *A modern EKOP model should not treat every high-count interval as news. It should ask whether order flow is abnormal after conditioning on the state of the market.*

References for this problem.

- Easley, Kiefer, O'Hara, and Paperman (1996), for the structural sequential-trade likelihood.
- Kokot (2004), for time-varying intensity extensions in sequential-trade econometrics.
- Cameron and Trivedi (2013), for count-data regression likelihoods.
- Lin and Ke (2011), for likelihood stability when intensities vary with large counts.

Problem 2.22 (Research: Hawkes flow toxicity versus VPIN toxicity)**.** Let buy and sell market-order arrivals be counted by N_t^B and N_t^S. Consider a bivariate Hawkes model

$$\lambda_t^B = \mu_B + \int_0^t \alpha_{BB} e^{-\beta(t-s)}\, dN_s^B + \int_0^t \alpha_{BS} e^{-\beta(t-s)}\, dN_s^S,$$

$$\lambda_t^S = \mu_S + \int_0^t \alpha_{SB} e^{-\beta(t-s)}\, dN_s^B + \int_0^t \alpha_{SS} e^{-\beta(t-s)}\, dN_s^S.$$

VPIN, in contrast, summarizes absolute imbalance in volume buckets.

(a) Interpret α_{SS} and α_{BS}.

(b) Derive the branching matrix and state the stability condition.

(c) Explain when VPIN and Hawkes toxicity should give similar signals.

(d) Explain when Hawkes toxicity contains information that VPIN discards.

Solution.

The parameter α_{SS} measures sell-to-sell self-excitation: a sell arrival raises the future sell intensity. Large α_{SS} can represent order splitting, panic selling, or liquidation cascades. The parameter α_{BS} measures the effect of a sell arrival on future buy intensity. It can represent replenishment, contrarian liquidity supply, or bid-side reaction after selling pressure.

For exponential kernels, the expected number of type-i offspring generated by one type-j event is α_{ij}/β. The branching matrix is

$$A = \frac{1}{\beta}\begin{pmatrix} \alpha_{BB} & \alpha_{BS} \\ \alpha_{SB} & \alpha_{SS} \end{pmatrix}.$$

The Hawkes process is stable when the spectral radius satisfies

$$\rho(A) < 1.$$

If $\rho(A)$ is close to one, shocks decay slowly and order flow can cluster strongly.

VPIN and Hawkes toxicity should give similar warnings when toxicity appears as persistent one-sided flow. For example, a sequence of sell-dominated buckets will raise VPIN and will also be consistent with large sell intensity in a Hawkes model.

Hawkes toxicity contains extra timing information. VPIN discards the order and spacing of events inside buckets. A Hawkes model distinguishes a smooth sell-heavy interval from a self-exciting burst in which each sell rapidly raises the chance of more sells. That distinction matters for crash dynamics, because clustered arrivals can overwhelm liquidity faster than the same total volume spread evenly through time. □

Takeaway. *VPIN is a bucket statistic; Hawkes toxicity is a dynamics model. The latter can see clustering and feedback that volume imbalance alone may miss.*

References for this problem.

- Hawkes (1971), for self-exciting point processes.
- Easley, Lopez de Prado, and O'Hara (2012), for VPIN.
- Bauwens and Hautsch (2009), for point-process modeling of financial high-frequency data.
- Bacry, Mastromatteo, and Muzy (2015), for Hawkes processes in finance.
- Andersen and Bondarenko (2014), for the VPIN critique.

Problem 2.23 (Research: Flash-crash warning signals and causal diagnostics)**.** Suppose an econometrician constructs a warning signal W_t from VPIN, Hawkes branching ratios, depth withdrawal, and recent volatility. A warning is issued when $W_t > c$. Let $C_{t,h} = 1$ denote that a flash-crash event occurs within the next horizon h.

(a) Define precision, recall, and the false-positive rate for the warning rule.

(b) Explain why in-sample fit is not enough for a crash-warning statistic.

(c) Propose an out-of-sample validation design.

(d) Explain why a successful warning signal is not automatically a causal explanation of the crash.

Solution.

Let

$$A_t(c) = 1\{W_t > c\}$$

be the alarm indicator. Precision is

$$\mathbb{P}(C_{t,h} = 1 \mid A_t(c) = 1),$$

the probability that an alarm is followed by a crash. Recall is

$$\mathbb{P}(A_t(c) = 1 \mid C_{t,h} = 1),$$

the probability that the rule catches crash episodes. The false-positive rate is

$$\mathbb{P}(A_t(c) = 1 \mid C_{t,h} = 0).$$

In-sample fit is weak evidence because crashes are rare and flexible warning rules can overfit a few dramatic episodes. A statistic may appear to rise before known crashes simply because thresholds, horizons, and transformations were chosen after seeing the event.

A disciplined validation design fixes the rule on a training period, chooses the threshold c without using the test crashes, and then evaluates precision, recall, and false positives on a later period or on different markets. The evaluation should include quiet periods, high-volatility non-crash periods, and known liquidity-demand episodes so that the rule is tested against hard near-misses.

Even if W_t predicts crashes, it need not cause them. VPIN, Hawkes self-excitation, and depth withdrawal may be symptoms of latent stress rather than drivers. A causal claim requires a source of exogenous variation, such as a market-structure change, a tick-size rule, a speed bump, an outage, or a credible instrument for liquidity withdrawal or order-flow toxicity. □

Takeaway. *The econometrics of crashes is not just prediction. A serious warning signal must survive out-of-sample testing and must be separated from causal claims.*

References for this problem.

- Easley, Lopez de Prado, and O'Hara (2012), for VPIN as a flow-toxicity warning measure.
- Andersen and Bondarenko (2014), for the critique of VPIN and the Flash Crash interpretation.
- Hasbrouck (2007), for empirical market microstructure measurement discipline.
- Hansen (1982) and Newey and West (1987), for moment evaluation and dependence-robust inference.

Bibliography

Andersen, Torben G., and Oleg Bondarenko (2014). "VPIN and the Flash Crash." *Journal of Financial Markets* 17, 1–46. [Critical reference for contested VPIN/flash-crash interpretation.]

Bacry, Emmanuel, Iacopo Mastromatteo, and Jean-Francois Muzy (2015). "Hawkes Processes in Finance." *Market Microstructure and Liquidity* 1(1), 1550005. [Survey extension for self-exciting order-flow models.]

Bauwens, Luc, and Nikolaus Hautsch (2009). "Modelling Financial High Frequency Data Using Point Processes." In *Handbook of Financial Time Series*, Springer, 953–979. [First-reading source for point-process econometrics.]

Cameron, A. Colin, and Pravin K. Trivedi (2013). *Regression Analysis of Count Data*. 2nd ed. Cambridge University Press. [Textbook source for Poisson and negative-binomial count likelihoods.]

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society: Series B* 39(1), 1–38. [Canonical EM reference for latent-regime likelihood estimation.]

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara (2002). "Is Information Risk a Determinant of Asset Returns?" *Journal of Finance* 57(5), 2185–2221. [Empirical PIN implementation and likelihood factorization reference.]

Easley, David, Nicholas M. Kiefer, and Maureen O'Hara (1997). "One Day in the Life of a Very Common Stock." *Review of Financial Studies* 10(3), 805–835. [Primary sequential-trade/PIN estimation source.]

Easley, David, Nicholas M. Kiefer, Maureen O'Hara, and Joseph B. Paperman (1996). "Liquidity, Information, and Infrequently Traded Stocks." *Journal of Finance* 51(4), 1405–1436. [Primary EKOP/PIN mixture likelihood source.]

Easley, David, Marcos M. Lopez de Prado, and Maureen O'Hara (2012). "Flow Toxicity and Liquidity in a High-Frequency World." *Review of Financial Studies* 25(5), 1457–1493. [Foundational VPIN and flow-toxicity source.]

Easley, David, and Maureen O'Hara (1987). "Price, Trade Size, and Information in Securities Markets." *Journal of Financial Economics* 19(1), 69–90. [Primary source for trade size as an information signal.]

Easley, David, and Maureen O'Hara (1992). "Time and the Process of Security Price Adjustment." *Journal of Finance* 47(2), 577–605. [Primary source for no-trade intervals as information.]

Engle, Robert F., and Jeffrey R. Russell (1998). "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data." *Econometrica* 66(5), 1127–1162. [Duration-model extension for transaction timing.]

Glosten, Lawrence R., and Lawrence E. Harris (1988). "Estimating the Components of the Bid/Ask Spread." *Journal of Financial Economics* 21(1), 123–142. [Empirical spread decomposition linked to trade size.]

Glosten, Lawrence R., and Paul R. Milgrom (1985). "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders." *Journal of Financial Economics* 14(1), 71–100. [Primary Bayesian adverse-selection quote-setting source.]

Hamilton, James D. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57(2), 357–384. [Canonical Markov-switching model source.]

Hamilton, James D. (1994). *Time Series Analysis.* Princeton University Press. [Textbook source for latent-state likelihoods and regime switching.]

Hansen, Lars Peter (1982). "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50(4), 1029–1054. [Canonical GMM and moment-testing source.]

Hasbrouck, Joel (2007). *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading.* Oxford University Press. [Main first-reading textbook for empirical microstructure.]

Hawkes, Alan G. (1971). "Spectra of Some Self-Exciting and Mutually Exciting Point Processes." *Biometrika* 58(1), 83–90. [Canonical Hawkes process source.]

Kokot, Stefan (2004). *The Econometrics of Sequential Trade Models: Theory and Applications Using High Frequency Data.* Lecture Notes in Economics and Mathematical Systems 538. Springer Berlin, Heidelberg. [Specialized source for sequential-trade econometric extensions.]

Lee, Charles M. C., and Mark J. Ready (1991). "Inferring Trade Direction from Intraday Data." *Journal of Finance* 46(2), 733–746. [Canonical trade-sign classification source.]

Lin, Haitao, and Wenjun Ke (2011). "A Computing Bias in Estimating the Probability of Informed Trading." *Journal of Financial Markets* 14(4), 625–640. [Primary source on numerical instability in PIN estimation.]

Madhavan, Ananth, Matthew Richardson, and Mark Roomans (1997). "Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks." *Review of Financial Studies* 10(4), 1035–1064. [Reduced-form bridge to structural belief updating.]

Newey, Whitney K., and Kenneth D. West (1987). "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3), 703–708. [Canonical HAC covariance source.]

O'Hara, Maureen (1995). *Market Microstructure Theory.* Blackwell. [Main theory textbook for

adverse selection and sequential trade.]

Chapter 3

Kyle-Style Equilibrium and Market Impact

This chapter moves from dealer quote-setting and sequential-trade models to strategic informed trading, market impact, and execution. Kyle's model is the spine: informed demand, noise trading, competitive market-maker pricing, and price impact are determined together in equilibrium. The later problems connect Kyle's λ to empirical price impact, transient impact, optimal execution, and metaorder evidence.

3.1 Kyle Equilibrium

Problem 3.1 (Core: Static Kyle equilibrium and linear pricing)**.** Let the terminal value of a risky asset be

$$v \sim N(0, \sigma_v^2),$$

and let noise-trader demand be

$$u \sim N(0, \sigma_u^2),$$

independent of v. A risk-neutral insider observes v and submits order $x = \beta v$. Competitive risk-neutral market makers observe only total order flow

$$y = x + u$$

and set the price

$$p = \mathbb{E}[v \mid y].$$

(a) If market makers conjecture $x = \beta v$, derive the linear pricing rule $p = \lambda y$.

(b) Given a linear pricing rule $p = \lambda y$, solve the insider's optimal order conditional on v.

(c) Solve for the linear equilibrium values of β and λ.

(d) Interpret market depth $1/\lambda$.

Solution.

Under the conjecture $x = \beta v$, total order flow is

$$y = \beta v + u.$$

Since (v, y) is jointly normal, the market maker's conditional expectation is linear:

$$p = \mathbb{E}[v \mid y] = \lambda y, \qquad \lambda = \frac{\mathrm{Cov}(v, y)}{\mathrm{Var}(y)}.$$

Now

$$\mathrm{Cov}(v, y) = \mathrm{Cov}(v, \beta v + u) = \beta \sigma_v^2$$

and

$$\mathrm{Var}(y) = \beta^2 \sigma_v^2 + \sigma_u^2.$$

Therefore

$$\lambda = \frac{\beta \sigma_v^2}{\beta^2 \sigma_v^2 + \sigma_u^2}.$$

Given a linear pricing rule, the insider chooses x to maximize expected profit conditional on v:

$$\mathbb{E}[(v - p)x \mid v] = \mathbb{E}[(v - \lambda(x + u))x \mid v] = x(v - \lambda x),$$

because $\mathbb{E}[u] = 0$. The first-order condition is

$$v - 2\lambda x = 0,$$

so

$$x = \frac{v}{2\lambda}.$$

In a linear equilibrium, this must equal βv, hence

$$\beta = \frac{1}{2\lambda}.$$

Combining the pricing equation with the insider's optimality condition gives

$$\frac{1}{2\beta} = \frac{\beta \sigma_v^2}{\beta^2 \sigma_v^2 + \sigma_u^2}.$$

Thus

$$\beta^2\sigma_v^2 + \sigma_u^2 = 2\beta^2\sigma_v^2,$$

so

$$\beta^2\sigma_v^2 = \sigma_u^2.$$

Taking the positive root,

$$\beta = \frac{\sigma_u}{\sigma_v}, \qquad \lambda = \frac{\sigma_v}{2\sigma_u}.$$

Market depth is the reciprocal of price impact:

$$\frac{1}{\lambda} = \frac{2\sigma_u}{\sigma_v}.$$

More noise trading makes the market deeper because the insider can hide more easily. More private-value uncertainty makes the market less deep because a given order flow is more informative about v. □

Takeaway. *Kyle's λ is not just a regression slope. It is an equilibrium object: the market maker's Bayesian updating rule and the insider's strategic demand must be mutually consistent.*

References for this problem.

- Kyle (1985), for the canonical linear insider-trading equilibrium.
- O'Hara (1995), for the textbook treatment of strategic informed trading.
- Hasbrouck (2007), for the empirical interpretation of price impact and market depth.

Problem 3.2 (Core: Comparative statics of Kyle's lambda)**.** Use the static Kyle equilibrium

$$\beta = \frac{\sigma_u}{\sigma_v}, \qquad \lambda = \frac{\sigma_v}{2\sigma_u}.$$

(a) How does λ change with private-value uncertainty σ_v?

(b) How does λ change with noise-trader volatility σ_u?

(c) How does insider aggressiveness β change with σ_v and σ_u?

(d) Compute $\mathrm{Var}(v \mid y)$ in equilibrium.

Solution.

The price-impact coefficient is

$$\lambda = \frac{\sigma_v}{2\sigma_u}.$$

Therefore

$$\frac{\partial \lambda}{\partial \sigma_v} = \frac{1}{2\sigma_u} > 0, \qquad \frac{\partial \lambda}{\partial \sigma_u} = -\frac{\sigma_v}{2\sigma_u^2} < 0.$$

Price impact is higher when the asset value is more uncertain, and lower when noise trading is larger.

The insider's trading coefficient is

$$\beta = \frac{\sigma_u}{\sigma_v}.$$

Hence insider demand is more aggressive when camouflage from noise trading is large, and less aggressive when the private signal is more valuable per unit of order flow.

The posterior variance is

$$\mathrm{Var}(v \mid y) = \mathrm{Var}(v) - \frac{\mathrm{Cov}(v, y)^2}{\mathrm{Var}(y)}.$$

In equilibrium,

$$\beta^2\sigma_v^2 = \sigma_u^2, \qquad \mathrm{Var}(y) = 2\sigma_u^2, \qquad \mathrm{Cov}(v, y) = \beta\sigma_v^2 = \sigma_u\sigma_v.$$

Therefore

$$\mathrm{Var}(v \mid y) = \sigma_v^2 - \frac{\sigma_u^2\sigma_v^2}{2\sigma_u^2} = \frac{\sigma_v^2}{2}.$$

□

Takeaway. *In the one-shot Gaussian Kyle auction, one observation of total order flow cuts the market maker's value uncertainty in half. Liquidity is the economic force that controls how quickly that information is revealed.*

References for this problem.

- Kyle (1985), for the equilibrium depth and information-revelation comparative statics.
- O'Hara (1995), for an accessible theoretical exposition.
- Hasbrouck (2007), for the link between theoretical depth and empirical price-impact measurement.

Problem 3.3 (Advanced: Multiple informed traders and Kyle depth)**.** Extend the one-period Kyle auction to $N \geq 1$ risk-neutral strategic traders. Each trader observes the same payoff v, noise demand is u, and

$$\mathrm{Var}(v) = \sigma_v^2, \qquad \mathrm{Var}(u) = \sigma_u^2, \qquad \mathrm{Cov}(u, v) = 0.$$

Assume a symmetric linear equilibrium in which trader n submits

$$x_n = bv,$$

so aggregate informed demand is

$$X = \sum_{n=1}^{N} x_n = Nbv.$$

The market maker observes

$$y = X + u$$

and sets a linear price $p = \lambda y$.

(a) Given the other $N-1$ traders' orders, derive trader n's best response to a linear price rule.

(b) Impose symmetry and solve for aggregate informed demand $X = \beta_N v$.

(c) Derive the corresponding price-impact coefficient λ_N.

(d) Explain how competition among informed traders changes market depth.

Solution.

Let the other traders submit aggregate order

$$X_{-n} = (N-1)bv.$$

If trader n chooses x, expected profit conditional on v is

$$x\,\mathbb{E}[v - \lambda(x + X_{-n} + u) \mid v] = x\{v - \lambda x - \lambda(N-1)bv\}.$$

The first-order condition is

$$v - \lambda(N-1)bv - 2\lambda x = 0.$$

In a symmetric linear equilibrium, $x = bv$, so

$$1 - \lambda(N-1)b - 2\lambda b = 0.$$

Thus

$$b = \frac{1}{\lambda(N+1)}.$$

Aggregate informed demand is therefore

$$X = Nbv = \beta_N v, \qquad \beta_N = Nb = \frac{N}{\lambda(N+1)}.$$

The market maker's projection coefficient is

$$\lambda = \frac{\mathrm{Cov}(v, y)}{\mathrm{Var}(y)} = \frac{\beta_N \sigma_v^2}{\beta_N^2 \sigma_v^2 + \sigma_u^2}.$$

Using

$$\beta_N = \frac{N}{\lambda(N+1)}$$

and solving gives

$$\beta_N = \sqrt{N}\,\frac{\sigma_u}{\sigma_v} \qquad \text{and} \qquad \lambda_N = \frac{\sqrt{N}}{N+1}\frac{\sigma_v}{\sigma_u}.$$

For $N = 1$, this reduces to the standard Kyle coefficient

$$\lambda_1 = \frac{\sigma_v}{2\sigma_u}.$$

As N grows, aggregate informed demand becomes more aggressive:

$$\beta_N = \sqrt{N}\,\frac{\sigma_u}{\sigma_v}.$$

But price impact is

$$\lambda_N = \frac{\sqrt{N}}{N+1}\frac{\sigma_v}{\sigma_u},$$

which eventually declines like $1/\sqrt{N}$. Competition among informed traders makes them reveal information more aggressively. The market becomes deeper in the sense that each unit of order flow carries less monopoly rents for any one insider, even though total informed trading is larger. □

Takeaway. *Kyle's monopoly-insider benchmark is only one point in a family. With many symmetrically informed traders, aggregate informed trading rises, individual market power falls, and the equilibrium price-impact coefficient becomes* $\lambda_N = \sqrt{N}\,\sigma_v/[(N+1)\sigma_u]$.

References for this problem.

- Noldeke and Troger (2006), for the multiple-strategic-trader Kyle model and the characterization of linear equilibria.
- Bagnoli, Viswanathan, and Holden (2001), for distributional assumptions behind linear equilibria with strategic traders.
- Kyle (1985), for the single-insider benchmark.

Problem 3.4 (Research: Distributional assumptions behind linear Kyle equilibria)**.** In the Gaussian Kyle model, market-maker pricing is linear because conditional expectations of jointly normal variables are linear. Noldeke and Troger (2006) show that the right broader distributional class is elliptical.

Suppose asset payoff v and noise demand u have finite second moments,

$$\mathrm{Var}(v) = \sigma_v^2 > 0, \qquad \mathrm{Var}(u) = \sigma_u^2 > 0, \qquad \mathrm{Cov}(u, v) = 0,$$

but they need not be independent. Let aggregate informed demand be linear,

$$X = \beta v,$$

so total order flow is

$$y = \beta v + u.$$

(a) State the condition needed for the market maker's pricing rule to be linear.

(b) Explain why joint normality is sufficient but not necessary.

(c) Explain why elliptical distributions preserve linear projection as conditional expectation.

(d) Explain the reverse lesson: why requiring linear equilibria for broad Kyle environments restricts the distributional family.

Solution.

The market maker prices competitively:

$$p(y) = \mathbb{E}[v \mid y].$$

Thus a linear pricing rule exists if

$$\mathbb{E}[v \mid \beta v + u]$$

is affine in $\beta v + u$. The linear projection coefficient is always

$$\frac{\mathrm{Cov}(v, y)}{\mathrm{Var}(y)} = \frac{\beta \sigma_v^2}{\beta^2 \sigma_v^2 + \sigma_u^2},$$

but the conditional expectation need not equal this projection unless the distribution has the right structure.

Joint normality is sufficient because normal conditional expectations are linear. But normality is stronger than needed. Elliptical distributions are affine transformations of spherically symmetric distributions. Their level sets are ellipses rather than arbitrary shapes. This symmetry implies that conditional means are linear functions of the conditioning variable whenever the relevant second moments exist.

Therefore, if (u, v) is elliptically distributed, the market maker's Bayesian pricing rule can remain linear even when u and v are not independent. This matters economically because noise trading may have a component related to fundamentals. A liquidity shock can be correlated with macro or firm conditions without destroying linear Kyle pricing, as long as the joint distribution has the elliptical structure.

The reverse direction is the deeper lesson. If one asks for linear equilibria to exist robustly across Kyle games with multiple symmetrically informed strategic traders, then the distribution cannot be arbitrary. Under moment determinacy, the requirement that conditional expectations behave linearly for the relevant order-flow combinations forces the standardized exogenous variables to have the rotation-invariance structure associated with elliptical families. Linear Kyle equilibria are therefore not just convenient algebra; they encode a strong distributional assumption. □

Takeaway. *Gaussianity is not the real knife-edge. The deeper condition behind linear Kyle pricing is ellipticity: enough symmetry for conditional means to be linear. When that symmetry fails, a linear price-impact coefficient may be only a projection, not the market maker's true Bayesian price.*

References for this problem.

- Noldeke and Troger (2006), for the characterization of distributions that imply linear equilibria in Kyle models.
- Foster and Viswanathan (1993), for Kyle-style models with correlated payoff and order-flow components.
- Bagnoli, Viswanathan, and Holden (2001), for the normal-distribution characterization under independence.

Problem 3.5 (Advanced: Finite market makers and the Kyle–Nash limit)**.** Kyle's competitive market-maker formulation can be viewed as the limit of a Nash game among finitely many market makers. Boulatov and Bernhardt (2015) study the Kyle (1983) version with $J \geq 3$ risk-neutral market makers.

Normalize

$$\sigma_v = \sigma_u = 1.$$

In the symmetric linear equilibrium, the insider trades

$$x = \beta^* v$$

and the inverse market depth is λ^*, with

$$\beta^* = \frac{J-2}{J-1}, \qquad \lambda^* = \frac{1}{2}\frac{J-1}{J-2}.$$

(a) Show that $\beta^* \lambda^* = 1/2$.

(b) Compute the limiting values of β^* and λ^* as $J \to \infty$.

(c) Compare the limit with the standard static Kyle equilibrium under $\sigma_v = \sigma_u = 1$.

(d) Explain economically why finite market-maker competition changes inverse market depth.

Solution.

Multiplying the two equilibrium coefficients gives

$$\beta^* \lambda^* = \frac{J-2}{J-1} \cdot \frac{1}{2} \frac{J-1}{J-2} = \frac{1}{2}.$$

Thus the insider's first-order condition has the same basic form as the standard Kyle condition:

$$x = \frac{v}{2\lambda^*}.$$

As the number of market makers grows,

$$\lim_{J \to \infty} \beta^* = \lim_{J \to \infty} \frac{J-2}{J-1} = 1$$

and

$$\lim_{J \to \infty} \lambda^* = \lim_{J \to \infty} \frac{1}{2} \frac{J-1}{J-2} = \frac{1}{2}.$$

The standard one-period Kyle equilibrium with $\sigma_v = \sigma_u = 1$ has

$$\beta = 1, \qquad \lambda = \frac{1}{2}.$$

So the finite-market-maker Nash model converges to the competitive Kyle pricing rule as J becomes large.

For finite J, the inverse depth is

$$\lambda^* = \frac{1}{2} \frac{J-1}{J-2} > \frac{1}{2}.$$

With fewer market makers, competition is weaker. Market makers shade supply schedules more aggressively, so the price response per unit of total order flow is larger. As J grows, competition erodes this strategic markup and the model approaches the competitive market-maker benchmark. □

Takeaway. *Competitive Kyle pricing can be seen as the large-J limit of a Nash game among market makers. Finite market-maker competition raises inverse depth, but the core insider optimality relation $\beta^* \lambda^* = 1/2$ remains.*

References for this problem.

- Boulatov and Bernhardt (2015), for robustness of the finite-market-maker Kyle–Nash equilibrium.

- Kyle (1983), for the finite-market-maker strategic supply-schedule formulation.
- Kyle (1985), for the competitive market-maker limiting benchmark.

Problem 3.6 (Research: Robustness of linear Kyle equilibrium to belief errors)**.** Boulatov and Bernhardt (2015) ask whether the linear Kyle equilibrium is robust when agents slightly misperceive others' strategies. Their parametric example keeps strategies linear but lets conjectures be slightly wrong.

Normalize $\sigma_v = \sigma_u = 1$. Suppose the insider actually uses the equilibrium intensity

$$\beta^* = \frac{J-2}{J-1},$$

but market makers conjecture β^c instead. In their calculation, each market maker's expected payoff is proportional to

$$\pi_M(\beta^c) = A\frac{\beta^c}{(\beta^c)^2 + (\beta^*)^2},$$

where $A > 0$ is constant. Similarly, suppose the insider conjectures inverse depth λ^c while actual inverse depth is λ^*. Let

$$\xi = \frac{\lambda^*}{\lambda^c}.$$

The insider's expected profit is proportional to

$$\pi_I(\xi) = \frac{1}{2\lambda^*}\xi(1 - \xi/2).$$

(a) Show that $\pi_M(\beta^c)$ is locally maximized at $\beta^c = \beta^*$.

(b) Show that the first-order payoff loss from a small market-maker belief error is zero.

(c) Show that $\pi_I(\xi)$ is locally maximized at $\xi = 1$.

(d) Interpret the robustness result economically.

Solution.

Differentiate the market-maker payoff function:

$$\pi_M(\beta^c) = A\frac{\beta^c}{(\beta^c)^2 + (\beta^*)^2}.$$

Then

$$\frac{d\pi_M}{d\beta^c} = A\frac{(\beta^c)^2 + (\beta^*)^2 - 2(\beta^c)^2}{[(\beta^c)^2 + (\beta^*)^2]^2}.$$

Thus

$$\frac{d\pi_M}{d\beta^c} = A\frac{(\beta^*)^2 - (\beta^c)^2}{[(\beta^c)^2 + (\beta^*)^2]^2}.$$

At $\beta^c = \beta^*$,

$$\left.\frac{d\pi_M}{d\beta^c}\right|_{\beta^c=\beta^*} = 0.$$

So a small conjectural error

$$\beta^c = \beta^* + \varepsilon$$

has no first-order effect on the market maker's expected payoff:

$$\pi_M(\beta^* + \varepsilon) - \pi_M(\beta^*) = O(\varepsilon^2).$$

For the insider,

$$\pi_I(\xi) = \frac{1}{2\lambda^*}\xi(1 - \xi/2).$$

Differentiating gives

$$\pi_I'(\xi) = \frac{1}{2\lambda^*}(1 - \xi).$$

Thus

$$\pi_I'(1) = 0,$$

so the payoff is locally maximized when the insider's conjecture is correct:

$$\lambda^c = \lambda^*.$$

If $\xi = 1 + \varepsilon$, then

$$\pi_I(1 + \varepsilon) - \pi_I(1) = O(\varepsilon^2).$$

This is the robustness idea. At the linear Kyle equilibrium, small errors in beliefs about the other side's linear strategy do not produce first-order payoff losses. The equilibrium is locally flat in those conjectural directions. That helps explain why the linear Kyle equilibrium is not merely a fragile exact solution: it remains locally stable to small strategic misperceptions. □

Takeaway. *Robustness means more than existence. In the Boulatov–Bernhardt sense, the linear Kyle equilibrium has zero first-order payoff loss from small belief errors, while nonlinear equilibria, if they exist, do not share this stability.*

References for this problem.

- Boulatov and Bernhardt (2015), for the robustness characterization of the linear Kyle–Nash equilibrium.
- Kyle (1983), for the finite-market-maker Nash formulation.
- Kyle (1985), for the competitive-market-maker limiting case.

Problem 3.7 (Research: Learning Kyle equilibrium with neural-network losses)**.** Friedrich and

Teichmann (2020) use neural networks to learn the one-period Kyle equilibrium from the agents' objectives rather than imposing the closed-form solution.

Let

$$Z \sim N(\mu_z, \sigma_z^2), \qquad Y \sim N(0, \sigma_y^2),$$

with Z and Y independent. The insider observes $Z = z$ and chooses an order

$$x = X_\theta(z),$$

where X_θ is a neural network. The market maker observes only

$$Q = X_\theta(Z) + Y$$

and prices with another neural network $P_\phi(Q)$.

(a) Write a squared-error loss whose minimizer is the market maker's efficient pricing rule.

(b) Write a loss for training the insider against a fixed pricing network.

(c) Explain why alternating the two training steps is an approximate best-response procedure.

(d) Explain why the Gaussian Kyle equilibrium is a useful benchmark before applying the method to less tractable variants.

Solution.

The market maker's equilibrium condition is

$$P(Q) = E[Z \mid Q].$$

The conditional expectation is the L^2 projection of Z on the information contained in Q. Thus a natural market-maker loss is

$$\mathcal{L}_M(\phi; \theta) = E\left[(Z - P_\phi(X_\theta(Z) + Y))^2\right].$$

For a fixed insider network X_θ, minimizing this loss over ϕ trains the pricing network to approximate the Bayesian price.

The insider wants to maximize expected trading profit,

$$E\left[X_\theta(Z)\,(Z - P_\phi(X_\theta(Z) + Y))\right].$$

Equivalently, one can train the insider by minimizing

$$\mathcal{L}_I(\theta; \phi) = -E\left[X_\theta(Z)\,(Z - P_\phi(X_\theta(Z) + Y))\right].$$

This is a differentiable version of the Kyle insider's objective: the insider chooses order size while internalizing the price impact generated by the market maker's pricing rule.

Alternating training has the form

$$\phi_{k+1} \approx \arg\min_{\phi} \mathcal{L}_M(\phi; \theta_k), \qquad \theta_{k+1} \approx \arg\min_{\theta} \mathcal{L}_I(\theta; \phi_{k+1}).$$

This is not merely curve fitting. It is an approximate best-response iteration: the market maker learns to price efficiently against the current insider, then the insider learns to trade optimally against the current market maker. A fixed point of these updates is a computational analogue of Kyle equilibrium.

Under Gaussian assumptions the analytic solution is known:

$$X(z) = \alpha + \beta z, \qquad P(q) = \mu + \lambda q,$$

where

$$\alpha = -\frac{\sigma_y}{\sigma_z}\mu_z, \qquad \beta = \frac{\sigma_y}{\sigma_z}, \qquad \mu = \mu_z, \qquad \lambda = \frac{\sigma_z}{2\sigma_y}.$$

If X is linear, then

$$Q = \alpha + \beta Z + Y,$$

so (Z, Q) is jointly normal and the conditional expectation $E[Z \mid Q]$ is linear. The market-maker squared-error problem therefore recovers the Kyle pricing rule. The insider's first-order condition against a linear price gives the same impact-trading balance as in the closed-form model.

This benchmark matters because it tests whether the computational environment can rediscover a known equilibrium. Once the neural agents learn the Gaussian case, the same machinery can be used to explore transaction costs, non-Gaussian values, constraints, or behavioral departures where closed-form Kyle algebra may no longer be available. □

Takeaway. *The neural Kyle model is not just a forecasting exercise. The market-maker network learns the conditional expectation $E[Z \mid Q]$, while the insider network learns a strategic best response to that pricing rule. In the Gaussian case, successful training should rediscover Kyle's linear equilibrium.*

References for this problem.

- Friedrich and Teichmann (2020), for neural-network learning of the single-period Kyle equilibrium.
- Kyle (1985), for the analytic Gaussian benchmark.
- Boulatov and Bernhardt (2015), for a complementary robustness view of the linear Kyle equilibrium.

Problem 3.8 (Research: Diagnosing a learned neural Kyle equilibrium)**.** Suppose the neural Kyle model is trained under the parameter values used as an illustrative non-centered case:

$$\mu_z = 0.5, \qquad \sigma_z = 2, \qquad \sigma_y = 1.$$

(a) Compute the theoretical Kyle coefficients $(\alpha, \beta, \mu, \lambda)$.

(b) Give two regression diagnostics for checking whether the trained insider and market maker have learned the linear equilibrium.

(c) Give one pricing-residual diagnostic for market-maker efficiency.

(d) Give one first-order-condition diagnostic for insider optimality when the learned price rule is approximately linear.

Solution.

The theoretical coefficients are

$$\alpha = -\frac{\sigma_y}{\sigma_z}\mu_z = -\frac{1}{2} \cdot 0.5 = -0.25,$$

$$\beta = \frac{\sigma_y}{\sigma_z} = \frac{1}{2},$$
$$\mu = \mu_z = 0.5,$$

and

$$\lambda = \frac{\sigma_z}{2\sigma_y} = \frac{2}{2} = 1.$$

Thus the benchmark equilibrium is

$$X(z) = -0.25 + 0.5z, \qquad P(q) = 0.5 + q.$$

A first diagnostic is to regress the learned insider order on the realized value:

$$X_\theta(Z) = a_X + b_X Z + \varepsilon_X.$$

Successful learning should give

$$a_X \approx -0.25, \qquad b_X \approx 0.5,$$

with little systematic nonlinearity in the residuals. A second diagnostic is to regress the learned market-maker price on total order flow:

$$P_\phi(Q) = a_P + b_P Q + \varepsilon_P.$$

Successful learning should give

$$a_P \approx 0.5, \qquad b_P \approx 1.$$

For market-maker efficiency, check whether the pricing residual has conditional mean zero:

$$E[Z - P_\phi(Q) \mid Q] \approx 0.$$

Empirically, this can be checked by binning observations by Q, regressing $Z - P_\phi(Q)$ on flexible functions of Q, or plotting residual means across order-flow quantiles. A learned price rule can look linear in a scatter plot but still fail this conditional-moment test.

For insider optimality, suppose the learned price rule is approximately

$$P_\phi(q) \approx \hat{\mu} + \hat{\lambda} q.$$

Against a linear price, the insider's conditional objective at value z is

$$x\left(z - \hat{\mu} - \hat{\lambda} x\right),$$

because $E[Y] = 0$. The first-order condition is therefore

$$z - \hat{\mu} - 2\hat{\lambda} x = 0.$$

A useful diagnostic is the residual

$$r_I(z) = z - \hat{\mu} - 2\hat{\lambda} X_\theta(z).$$

If the insider network has learned a best response, this residual should be small across the high-probability range of Z. □

Takeaway. *Learning an equilibrium requires more than a low training loss. The trained agents should pass economic diagnostics: the insider's order should match Kyle's trading intensity, the market maker's price should be conditionally efficient, and the insider's first-order-condition residual should be small.*

References for this problem.

- Friedrich and Teichmann (2020), for neural-network implementation and training diagnostics in the one-period Kyle model.
- Kyle (1985), for the closed-form coefficients used as benchmark diagnostics.
- Hasbrouck (1991), for the broader empirical habit of diagnosing price impact through information content and price response.

Problem 3.9 (Research: Stochastic liquidity and liquidity timing)**.** Collin-Dufresne and Fos (2016) extend continuous-time Kyle by allowing the volatility of uninformed trading to be stochastic. Let σ_t^u denote the instantaneous volatility of noise trading. The terminal value v is fixed and known to the insider, but not to market makers.

Suppose the equilibrium price impact λ_t is lower in high-liquidity states, and the insider trades at an intensity proportional to the current mispricing:

$$\theta_t = \kappa_t(v - P_t).$$

(a) Explain why the insider wants κ_t to be larger when σ_t^u is high and λ_t is low.

(b) Explain why stochastic liquidity creates an option value of waiting.

(c) Explain why equilibrium price impact can be expected to rise on average over time.

(d) Contrast this with the standard Kyle model with constant noise-trading volatility.

Solution.

The insider earns more from a unit of informed trading when the market maker's price response is smaller. If liquidity trading is more volatile, the insider can hide more easily inside uninformed order flow. A high σ_t^u state therefore provides more camouflage. If, in the same state, equilibrium price impact λ_t is low, then a given informed order moves the price less against the insider. Both forces push toward a larger trading intensity:

$$\sigma_t^u \uparrow, \quad \lambda_t \downarrow \quad \Longrightarrow \quad \kappa_t \uparrow.$$

Economically, the insider times liquidity: he trades more aggressively when uninformed volume is abundant and market depth is high.

Stochastic liquidity changes the dynamic problem because the insider no longer faces the same trading conditions at all dates. If liquidity is poor today, the insider can wait for a future state in which noise trading is higher and price impact is lower. This option to wait is absent from the constant-liquidity Kyle benchmark. It is valuable because the insider's information is long lived: the terminal value v does not disappear merely because the insider delays some trading.

In equilibrium, however, market makers understand this timing option. If price impact were too low early on, the insider would prefer to wait too much for future high-liquidity states. To induce information revelation before maturity, equilibrium price impact must compensate for the insider's option to delay. Collin-Dufresne and Fos show that this can make price impact a submartingale:

$$E_t[\lambda_s] \geq \lambda_t, \qquad s > t.$$

That is, price impact is expected to rise on average.

This is very different from the basic Kyle model. With constant noise-trading volatility, liquidity conditions are stationary, Kyle's lambda is constant, and the insider reveals information at a smooth deterministic rate. With stochastic liquidity, the rate of informed trading, price impact, and price discovery all become state dependent. □

Takeaway. *Stochastic liquidity turns Kyle trading into a timing problem. The insider does not only ask how much private information to reveal; he also asks when liquidity is deep enough to reveal it cheaply.*

References for this problem.

- Collin-Dufresne and Fos (2016), for continuous-time Kyle equilibrium with stochastic noise-trading volatility.
- Back (1992), for the continuous-time Kyle bridge benchmark.
- Back and Pedersen (1998), for deterministic time variation in noise trading volatility.

Problem 3.10 (Research: Endogenous volume-volatility in stochastic Kyle)**.** In standard Kyle, price volatility is tied to uncertainty about the terminal fundamental. Collin-Dufresne and Fos (2016) show that stochastic uninformed volume can generate stochastic price volatility even when the terminal value is fixed and no new fundamental news arrives.

Let L_t denote a liquidity state driven by uninformed trading volatility. A reduced-form way to summarize the equilibrium price process is

$$dP_t = \sigma_P(L_t, P_t)\, dB_t,$$

with

$$\sigma_P(L_t, P_t)^2 = a(L_t)\, \Sigma_t,$$

where $\Sigma_t = \mathrm{Var}_t(v)$ is the remaining posterior uncertainty about the terminal value and $a(L_t)$ is the endogenous rate of information revelation.

(a) Explain why $a(L_t)$ should be higher in high uninformed-volume states.

(b) Show how this creates a positive relation between volume and price volatility.

(c) Interpret the price process as a subordinated or time-changed process.

(d) Explain why average price impact can be a poor proxy for realized execution costs in this model.

Solution.

When uninformed volume is high, the insider has more camouflage. Market makers also know this. In equilibrium, the insider trades more aggressively in those states, so private information is incorporated faster into prices. Thus the information-revelation rate satisfies the qualitative relation

$$L_t \uparrow \quad \Longrightarrow \quad a(L_t) \uparrow .$$

The conditional instantaneous variance of price changes is

$$\mathrm{Var}_t(dP_t) = \sigma_P(L_t, P_t)^2 \, dt = a(L_t)\Sigma_t \, dt.$$

Holding posterior uncertainty Σ_t fixed, a higher liquidity state raises $a(L_t)$ and therefore raises price volatility. This produces a positive volume-volatility relation even though the terminal fundamental has not become more volatile. The volatility comes from strategic information revelation, not from new payoff news.

The same idea can be written as an endogenous time change. Define the information clock

$$A_t = \int_0^t a(L_s) \, ds.$$

Then price discovery runs faster when A_t accumulates faster. High-volume states accelerate the clock; low-volume states slow it down. Prices therefore look like a bridge to the terminal value under a stochastic clock driven by uninformed volume. This links the model to mixture-of-distributions and subordinated-price views, but here the time change is generated endogenously by strategic trading.

Execution costs are also path dependent. Liquidity traders do not simply pay a cost determined by the calendar-time average of λ_t. They pay costs when their orders arrive, and those orders arrive precisely in states where noise-trading volatility, insider aggressiveness, and price discovery may all be different. Therefore

$$\frac{1}{T} \int_0^T \lambda_t \, dt$$

can be a poor summary of realized slippage. What matters is the joint path of uninformed volume, price impact, and insider trading intensity. □

Takeaway. *In stochastic-liquidity Kyle, volume is not a passive scale variable. Because high uninformed volume changes the insider's optimal timing, it endogenously changes price volatility, price discovery, and realized execution costs.*

References for this problem.

- Collin-Dufresne and Fos (2016), for the endogenous stochastic volume-volatility relation in Kyle equilibrium.

- Clark (1973), for subordinated price processes.
- Lamoureux and Lastrapes (1990), for the empirical link between volume and conditional heteroskedasticity.
- Gallant, Rossi, and Tauchen (1992), for empirical evidence on the price-volume-volatility relation.

Problem 3.11 (Advanced: Kyle's lambda as an identification problem)**.** Suppose an econometrician observes signed order flow q_t and price changes Δp_t and estimates

$$\Delta p_t = \theta q_t + e_t.$$

In the structural Kyle model, however, total order flow is

$$q_t = x_t + u_t,$$

where x_t is informed demand and u_t is liquidity demand.

(a) Write the population OLS coefficient θ.

(b) Explain when θ can be interpreted as Kyle's λ.

(c) Explain why predictable order flow creates an identification problem.

(d) State a moment condition based on order-flow innovations.

Solution.

The population slope is

$$\theta = \frac{\mathrm{Cov}(\Delta p_t, q_t)}{\mathrm{Var}(q_t)}.$$

This equals Kyle's λ only under a tight timing and information condition: the price change must be the market maker's response to the same unexpected total order flow observed by the market maker, and the residual must be orthogonal to that order-flow innovation.

In real high-frequency data, raw signed order flow is persistent. Some trades are predictable from lagged order flow, inventory pressure, execution algorithms, or intraday liquidity demand. A regression of price changes on raw q_t therefore mixes several objects:

$$\text{information response} \quad + \quad \text{mechanical liquidity demand} \quad + \quad \text{predictable execution pressure.}$$

The slope is an average response to signed volume, not automatically the structural price impact of unexpected informed order flow.

A cleaner empirical target uses the order-flow innovation. Let

$$\nu_t = q_t - \mathbb{E}[q_t \mid \mathcal{F}_{t-1}]$$

be the unpredictable component of order flow. A Kyle-style moment condition is

$$\mathbb{E}\left[\nu_t \left(\Delta p_t - \lambda \nu_t\right)\right] = 0.$$

Equivalently,

$$\lambda = \frac{\mathrm{Cov}(\Delta p_t, \nu_t)}{\mathrm{Var}(\nu_t)}.$$

This does not solve every endogeneity problem, but it respects the central Kyle idea: prices respond to the information in order-flow surprises, not to the predictable part of trade splitting. □

Takeaway. *Estimating price impact is not the same as running price changes on signed volume. The econometric habit is to ask: impact of which component of order flow, relative to which information set?*

References for this problem.

- Kyle (1985), for structural λ as an equilibrium pricing rule.
- Hasbrouck (1991), for measuring the information content and ultimate price impact of trade innovations.
- Hasbrouck (2007), for the empirical market-microstructure treatment of impact regressions and order-flow innovations.

Problem 3.12 (Advanced: Dynamic Kyle and gradual revelation)**.** Consider a discrete-time Kyle-style learning problem. At the start of auction n, the market maker's prior uncertainty about v is

$$\mathrm{Var}(v \mid \mathcal{F}_{n-1}) = \Sigma_{n-1}.$$

The insider submits

$$x_n = \beta_n (v - p_{n-1}),$$

noise demand is $u_n \sim N(0, \sigma_u^2)$, and the market maker observes

$$y_n = x_n + u_n.$$

The price update is linear:

$$p_n = p_{n-1} + \lambda_n y_n.$$

(a) Derive λ_n as a projection coefficient.

(b) Derive the posterior variance Σ_n after observing y_n.

(c) Explain why dynamic Kyle models generate gradual information revelation.

Solution.

Conditional on $\mathcal{F}_{n-1}$, the relevant unknown is

$$z_n = v - p_{n-1}, \qquad \mathrm{Var}(z_n \mid \mathcal{F}_{n-1}) = \Sigma_{n-1}.$$

Order flow is

$$y_n = \beta_n z_n + u_n.$$

The market maker's linear projection coefficient is

$$\lambda_n = \frac{\mathrm{Cov}(z_n, y_n \mid \mathcal{F}_{n-1})}{\mathrm{Var}(y_n \mid \mathcal{F}_{n-1})} = \frac{\beta_n \Sigma_{n-1}}{\beta_n^2 \Sigma_{n-1} + \sigma_u^2}.$$

The posterior variance after observing y_n is the usual Gaussian projection variance:

$$\Sigma_n = \Sigma_{n-1} - \frac{\mathrm{Cov}(z_n, y_n \mid \mathcal{F}_{n-1})^2}{\mathrm{Var}(y_n \mid \mathcal{F}_{n-1})}.$$

Substituting the covariance and variance gives

$$\Sigma_n = \Sigma_{n-1} - \frac{\beta_n^2 \Sigma_{n-1}^2}{\beta_n^2 \Sigma_{n-1} + \sigma_u^2}.$$

Equivalently, using the expression for λ_n,

$$\Sigma_n = \Sigma_{n-1} - \lambda_n \beta_n \Sigma_{n-1}.$$

The insider does not generally reveal all information immediately because a large order moves price against him. Noise trading gives camouflage, but only gradually. Dynamic Kyle equilibria trade off the value of exploiting current private information against the cost of making future trades less profitable by revealing too much today. □

Takeaway. *Dynamic Kyle turns price impact into a filtering problem. Each auction reveals part of the insider's information, and the remaining posterior variance governs future trading aggressiveness.*

References for this problem.

- Kyle (1985), for sequential auctions and gradual information revelation.
- Back (1992), for the continuous-time Kyle model.

- O'Hara (1995), for the theory-textbook treatment of dynamic strategic trading.

Problem 3.13 (Advanced: Public disclosure of insider trades as a second signal)**.** In the standard dynamic Kyle model, the market maker observes total order flow during trading but not the insider's order separately. Huddart, Hughes, and Levine (2001) study what changes when the insider's trade is publicly disclosed after the trading round.

Let

$$z = v - p_{n-1}, \qquad z \sim N(0, \Sigma),$$

be the insider's remaining information at the start of a period. Suppose the insider's order is

$$x = az + \eta,$$

where $\eta \sim N(0, \tau^2)$ is an independent dissimulation term chosen by the insider. Liquidity demand is $u \sim N(0, \sigma_u^2)$, so the market maker first observes

$$y = x + u.$$

After the round, the insider's order x itself is publicly disclosed.

(a) Compute the posterior variance of z after observing total order flow y.

(b) Compute the posterior variance of z after the disclosed insider trade x is observed.

(c) Show what happens if the insider adds no dissimulation noise ($\tau^2 = 0$).

(d) Explain why disclosure accelerates price discovery and lowers future insider profits.

Solution.

Because

$$y = az + \eta + u,$$

the observation noise in y is $\eta + u$, with variance

$$\tau^2 + \sigma_u^2.$$

The Gaussian posterior variance after observing y is

$$\mathrm{Var}(z \mid y) = \Sigma - \frac{\mathrm{Cov}(z, y)^2}{\mathrm{Var}(y)}.$$

Now

$$\mathrm{Cov}(z, y) = a\Sigma \qquad \text{and} \qquad \mathrm{Var}(y) = a^2\Sigma + \tau^2 + \sigma_u^2.$$

Thus

$$\mathrm{Var}(z \mid y) = \Sigma - \frac{a^2\Sigma^2}{a^2\Sigma + \tau^2 + \sigma_u^2} = \Sigma \frac{\tau^2 + \sigma_u^2}{a^2\Sigma + \tau^2 + \sigma_u^2}.$$

After disclosure, the market maker observes

$$x = az + \eta.$$

The posterior variance based on the disclosed insider trade is

$$\mathrm{Var}(z \mid x) = \Sigma - \frac{\mathrm{Cov}(z, x)^2}{\mathrm{Var}(x)}.$$

Since

$$\mathrm{Cov}(z, x) = a\Sigma, \qquad \mathrm{Var}(x) = a^2\Sigma + \tau^2,$$

we get

$$\mathrm{Var}(z \mid x) = \Sigma \frac{\tau^2}{a^2\Sigma + \tau^2}.$$

If $\tau^2 = 0$, then $x = az$ perfectly reveals z whenever $a \neq 0$, and

$$\mathrm{Var}(z \mid x) = 0.$$

So without dissimulation, the first disclosed trade destroys the insider's remaining informational advantage for future rounds.

Disclosure therefore changes the dynamic trade-off. A larger current informed trade earns more immediate profit, but it also makes future profits smaller because the public record reveals information. The insider responds by adding noise to his own trade. This slows the inference from the disclosed trade, but it also makes current trading less directly aligned with the private signal. □

Takeaway. *Trade disclosure turns the insider's own order into a public signal. Dynamic Kyle then has two information channels: order flow during the auction and the disclosed insider order after the auction.*

References for this problem.

- Huddart, Hughes, and Levine (2001), for the Kyle model with ex-post public disclosure of insider trades.
- Kyle (1985), for the no-disclosure dynamic benchmark.
- O'Hara (1995), for the broader theory background on insider trading and disclosure.

Problem 3.14 (Research: Dissimulation and contrarian insider trades)**.** Continue the disclosure setting. The insider uses

$$x = az + \eta, \qquad z = v - p_{n-1},$$

where η is independent noise added by the insider. Huddart, Hughes, and Levine call this behavior dissimulation: the insider garbles his own trade so that disclosure reveals less information.

Assume the insider chooses the dissimulation variance so that

$$\tau^2 = \mathrm{Var}(\eta) = a^2\Sigma, \qquad \Sigma = \mathrm{Var}(z).$$

(a) Compute $\mathrm{Var}(x)$.

(b) Compute $\mathrm{Var}(z \mid x)$ under this variance choice.

(c) Show that the insider sometimes trades against his information.

(d) If the insider also wants $\mathrm{Var}(x) = \sigma_u^2$, solve for a.

Solution.

The insider's order has two independent components:

$$x = az + \eta.$$

Therefore

$$\mathrm{Var}(x) = a^2\Sigma + \tau^2.$$

Under the dissimulation choice $\tau^2 = a^2\Sigma$,

$$\mathrm{Var}(x) = 2a^2\Sigma.$$

From the previous problem,

$$\mathrm{Var}(z \mid x) = \Sigma\frac{\tau^2}{a^2\Sigma + \tau^2}.$$

Substituting $\tau^2 = a^2\Sigma$ gives

$$\mathrm{Var}(z \mid x) = \frac{\Sigma}{2}.$$

So the disclosed insider trade reveals exactly half of the remaining variance in this simplified Gaussian calculation. The insider has not hidden completely, but he has avoided full revelation.

Now consider the sign of the trade. If $z > 0$, the information-based component az is positive, but the realized trade is

$$x = az + \eta.$$

The insider sells, despite good news, whenever

$$x < 0 \iff \eta < -az.$$

Because η is Gaussian with full support,

$$\mathbb{P}(x < 0 \mid z > 0) = \Phi\left(-\frac{az}{\tau}\right) > 0.$$

Similarly, if $z < 0$, the insider sometimes buys despite bad news. These are contrarian trades, but they are not manipulation in the sense of pretending to have the opposite signal. They arise because the insider deliberately adds noise to slow down future inference from disclosure.

If the insider also wants the variance of his own order to match liquidity demand,

$$\mathrm{Var}(x) = \sigma_u^2,$$

then

$$2a^2\Sigma = \sigma_u^2.$$

Thus

$$a = \frac{\sigma_u}{\sqrt{2\Sigma}},$$

taking the positive root for informed buying on positive information. □

Takeaway. *Dissimulation gives a clean rationale for contrarian insider trades. A trade in the wrong direction can be part of an optimal disclosure-management strategy, because the insider is managing what future observers learn from the public record.*

References for this problem.

- Huddart, Hughes, and Levine (2001), for dissimulation, public disclosure, and contrarian insider trades.
- Kyle (1985), for the benchmark without post-trade disclosure.
- Fishman and Hagerty (1995), for contrast with manipulation-based explanations of misleading trades.

Problem 3.15 (Research: Information sharing and trading against error)**.** Consider a Kyle-style market with two informed investors. Investor H observes the fundamental value v perfectly. Investor L observes a noisy signal

$$y = v + e, \qquad e \sim N(0, \rho^{-1}),$$

where e is independent of v. Suppose L trades linearly on his signal,

$$x_L = by = bv + be, \qquad b > 0.$$

If L shares y with H, then H observes both v and $e = y - v$. The market maker observes total order flow

$$\omega = x_L + x_H + u$$

and prices linearly, $p = \lambda\omega$.

(a) Decompose L's order into a fundamental component and a signal-error component.

(b) Given x_L, derive H's optimal order x_H conditional on (v, e).

(c) Show that H trades against the error component in L's signal.

(d) Explain why this can make the less informed investor willing to share information with the more informed investor.

Solution.

Investor L's order is

$$x_L = b(v + e) = bv + be.$$

The term bv is the part aligned with the fundamental. The term be is the part generated by signal error. If $e > 0$, investor L overestimates the fundamental and buys too much; if $e < 0$, he underestimates it and sells too much or buys too little.

Given the pricing rule $p = \lambda(x_L + x_H + u)$ and $\mathbb{E}[u] = 0$, investor H's expected profit conditional on (v, e) is

$$\mathbb{E}[x_H(v - p) \mid v, e] = x_H\{v - \lambda(x_L + x_H)\}.$$

The first-order condition is

$$v - \lambda x_L - 2\lambda x_H = 0.$$

Thus

$$x_H^* = \frac{v - \lambda x_L}{2\lambda} = \frac{v}{2\lambda} - \frac{x_L}{2}.$$

Substituting $x_L = bv + be$ gives

$$x_H^* = \left(\frac{1}{2\lambda} - \frac{b}{2}\right) v - \frac{b}{2} e.$$

The coefficient on the signal error e is negative:

$$-\frac{b}{2} < 0.$$

So, after seeing L's signal, H trades against the part of L's order flow that is due to error. If L buys too much because $e > 0$, H reduces demand or sells against him. If L sells too much because $e < 0$, H buys against him. This is the trading-against-error effect.

From L's perspective, this can be useful. His own noisy order creates price impact against him. By sharing his signal with a better-informed investor, he invites that investor to offset the error-driven part of his trade. That lowers the price impact of his mistake. The paradox is that information does not flow from the better-informed investor to the worse-informed investor. It can flow from the coarsely informed investor to the well-informed investor because the receiver can trade against the sender's signal error. □

Takeaway. *Information sharing can be strategic price-impact management. A less informed trader may reveal his signal not because it is high-quality information, but because a better-informed trader can absorb the error component and reduce the sender's own price impact.*

References for this problem.

- Goldstein, Xiong, and Yang (2025), for endogenous information sharing in a Kyle market and the trading-against-error mechanism.
- Kyle (1985), for the underlying strategic trading and linear pricing framework.
- O'Hara (1995), for the textbook background on informed trading and market-maker price setting.

Problem 3.16 (Advanced: Imperfect information and the CARA risk penalty)**.** In the standard risk-neutral Kyle model, an insider with an imperfect signal can often replace the terminal fundamental by its conditional expectation. This shortcut fails under risk aversion. To see why, suppose the terminal fundamental is

$$V_f = \Xi - G,$$

where the insider observes Ξ but not G, and

$$G \sim N(0, \varepsilon).$$

Let Q be the insider's final asset position and let P be the execution price, treated here as fixed conditional on the insider's information. Terminal trading wealth from the terminal mismatch is

$$W = Q(V_f - P) = Q(\Xi - P) - QG.$$

The insider has CARA utility with risk-aversion parameter $\gamma > 0$.

(a) Compute the certainty equivalent of W conditional on Ξ and P.

(b) Show that the residual uncertainty G creates a quadratic inventory penalty.

(c) Explain why the same residual uncertainty would not change a risk-neutral insider's objective.

(d) Interpret the penalty as the central economic difference between imperfect information with and without risk aversion.

Solution.

Conditional on Ξ and P, the only remaining random variable in wealth is G. Since G is Gaussian,

$$W = Q(\Xi - P) - QG$$

is Gaussian with conditional mean

$$\mathbb{E}[W \mid \Xi, P] = Q(\Xi - P)$$

and conditional variance

$$\mathrm{Var}(W \mid \Xi, P) = Q^2\varepsilon.$$

For CARA utility and Gaussian wealth, the certainty equivalent is mean minus one-half risk aversion times variance:

$$CE(W \mid \Xi, P) = Q(\Xi - P) - \frac{\gamma}{2}\varepsilon Q^2.$$

Equivalently, the paper writes the penalty through

$$g(Q) = \frac{1}{\gamma}\log \mathbb{E}\left[e^{\gamma GQ}\right].$$

Because $G \sim N(0, \varepsilon)$,

$$\mathbb{E}\left[e^{\gamma GQ}\right] = \exp\left(\frac{\gamma^2\varepsilon Q^2}{2}\right),$$

so

$$g(Q) = \frac{\gamma\varepsilon Q^2}{2}.$$

Thus the risk-adjusted value of the position is

$$Q(\Xi - P) - g(Q) = Q(\Xi - P) - \frac{\gamma\varepsilon Q^2}{2}.$$

A risk-neutral insider maximizes expected wealth. Since $\mathbb{E}[G] = 0$, the unobserved residual G drops out of the objective:

$$\mathbb{E}[W \mid \Xi, P] = Q(\Xi - P).$$

The insider can behave as if $\Xi = \mathbb{E}[V_f \mid \Xi]$ were the relevant fundamental. A risk-averse insider cannot do this, because the final position Q loads on the residual terminal jump $-G$. □

Takeaway. *Imperfect information is innocuous in the risk-neutral Kyle model because only conditional means matter. With CARA utility, the unlearned terminal component becomes inventory risk, producing the quadratic penalty* $\gamma\varepsilon Q^2/2$.

References for this problem.

- Chhaibi, Ekren, and Noh (2025), for the Gaussian Kyle model with imperfect information and risk aversion.
- Kyle (1985), for the risk-neutral benchmark where conditional expectations summarize private information.
- Back (1992), for the continuous-time Kyle benchmark.

Problem 3.17 (Advanced: Risk aversion dampens Kyle-style informed demand)**.** Continue the previous setup. In a one-period Kyle-style approximation, suppose the insider starts with random inventory β and chooses a trade x, so the final position is

$$Q = x + \beta.$$

Let the price be linear in the insider's order,

$$P = \lambda(x + u),$$

where $\lambda > 0$ and $\mathbb{E}[u] = 0$. Conditional on Ξ and β, approximate the insider's objective by the CARA certainty equivalent

$$J(x) = (x + \beta)(\Xi - \lambda x) - \frac{\gamma\varepsilon}{2}(x + \beta)^2.$$

(a) Derive the insider's optimal trade x^*.

(b) Compare it with the risk-neutral, perfectly informed benchmark.

(c) Show how residual terminal uncertainty ε reduces trading aggressiveness.

(d) Explain the equilibrium implication for price impact and volatility.

Solution.

Differentiate the objective:

$$J'(x) = \Xi - \lambda x - \lambda(x + \beta) - \gamma\varepsilon(x + \beta).$$

Therefore the first-order condition is

$$\Xi - 2\lambda x - \lambda\beta - \gamma\varepsilon x - \gamma\varepsilon\beta = 0.$$

Collecting terms in x gives

$$(2\lambda + \gamma\varepsilon)x = \Xi - (\lambda + \gamma\varepsilon)\beta.$$

Thus

$$x^* = \frac{\Xi - (\lambda + \gamma\varepsilon)\beta}{2\lambda + \gamma\varepsilon}.$$

If $\beta = 0$, this simplifies to

$$x^* = \frac{\Xi}{2\lambda + \gamma\varepsilon}.$$

The risk-neutral benchmark is obtained by setting $\gamma\varepsilon = 0$:

$$x^{RN} = \frac{\Xi}{2\lambda}.$$

Since

$$2\lambda + \gamma\varepsilon > 2\lambda,$$

the risk-averse imperfectly informed insider trades less aggressively on the same observed signal Ξ.

The term $\gamma\varepsilon$ is the product of risk aversion and residual fundamental uncertainty. Higher ε means the insider is less sure about the final payoff even after observing Ξ. Higher γ means this remaining uncertainty is more costly. Both forces reduce the signal loading in the insider's demand.

In equilibrium, less aggressive informed demand makes total order flow less informative about the terminal fundamental. The market maker therefore has less reason to move prices sharply in response to demand. This is the intuition behind the paper's result that greater uncertainty about the terminal price can make the market maker less reactive and the price process less volatile. □

Takeaway. *Risk aversion converts imperfect information into a force that restrains informed trading. The insider still trades on the signal, but the denominator contains a new risk term, $\gamma\varepsilon$, that has no role in the risk-neutral Kyle benchmark.*

References for this problem.

- Chhaibi, Ekren, and Noh (2025), for solvability of the Gaussian Kyle model with imperfect information and risk aversion.
- Kyle (1985), for the one-period linear benchmark.
- Back (1992), for the continuous-time Kyle model that the new paper extends.

3.2 Discrete Kyle Games and Inconspicuous Trading

Problem 3.18 (Advanced: Finite-action Kyle games and equilibrium existence)**.** The Gaussian Kyle model is elegant because linearity gives a closed-form equilibrium. Lorenz (2024) studies a different question: when does a Kyle equilibrium exist if the value distribution, insider information, and noise trader demand are not Gaussian?

Consider a finite-horizon Kyle game with dates $t = 1, \dots, T$. The asset value v takes values in a finite set, the insider receives private information over time, the noise trader's orders take values in finite sets, and the insider's admissible orders also come from finite sets. At each date, the market maker observes only total order flow and sets

$$p_t = \mathbb{E}[v \mid \mathcal{F}_t^Y],$$

where $\mathcal{F}_t^Y$ is the public history generated by total order flow.

(a) Explain why the insider's dynamic strategy set is finite-dimensional and compact after mixed strategies are allowed.

(b) Explain why the market maker's pricing rule can be written as a conditional-expectation response to the insider's strategy.

(c) State the fixed-point logic behind existence of equilibrium.

(d) Explain why this existence argument does not need Gaussianity or linear strategies.

Solution.

With finitely many dates, finitely many private histories, and finitely many orders at each private history, an insider strategy is just a list of probability vectors. Each vector lies in a simplex. The full mixed strategy space is a finite product of simplexes, hence compact and convex.

Given an insider strategy and the exogenous distribution of noise orders, every public order-flow history has an induced probability. The market maker's pricing rule is then determined by Bayesian updating:

$$p_t(y_1, \dots, y_t) = \mathbb{E}[v \mid Y_1 = y_1, \dots, Y_t = y_t],$$

with arbitrary choices only at histories that have zero probability under the candidate strategy. Thus, for any candidate insider strategy, the market maker has a rational pricing response.

Given the pricing response, the insider faces a finite dynamic optimization problem. At each private history, continuation payoffs are finite sums over future values, noise orders, and prices. A best response therefore exists. An equilibrium is a fixed point in which the insider strategy is optimal given the pricing rule, and the pricing rule is Bayesian given the induced distribution of total order flow.

The key point is that existence is topological rather than algebraic. Gaussian Kyle models use normality to guess and verify linear policies. A finite-action Kyle game can use mixed strategies, compactness, and continuity of expected payoffs to obtain an equilibrium without assuming normal fundamentals, normal noise demand, or a linear market-maker rule. □

Takeaway. *Closed-form Kyle equilibria are special, but existence is more robust. Once the dynamic game is finite, equilibrium can be built from compact strategy spaces, Bayesian pricing, and fixed-point logic rather than from Gaussian-linear structure.*

References for this problem.

- Lorenz (2024), for finite-action sequential Kyle games and equilibrium existence without probabilistic restrictions.
- Kyle (1985), for the original Gaussian-linear benchmark.
- Kreps and Wilson (1982), for sequential equilibrium in dynamic games with incomplete information.

Problem 3.19 (Research: Inconspicuous trading as a binomial bridge)**.** In an inconspicuous Kyle equilibrium, the market maker cannot detect the insider's presence from the distribution of total demand. Lorenz (2024) studies this idea in a binomial Kyle model. Let the noise trader's demand be a simple symmetric random walk

$$Z_n = \sum_{k=1}^{n} \zeta_k, \qquad \mathbb{P}(\zeta_k = 1) = \mathbb{P}(\zeta_k = -1) = \frac{1}{2}.$$

Let total demand be $Y_n = X_n + Z_n$, where X_n is the insider's cumulative demand. The inconspicuousness requirement is

$$(Y_0, \ldots, Y_N) \overset{d}{=} (Z_0, \ldots, Z_N)$$

from the market maker's perspective.

Suppose the fundamental has two values. A high value should make the insider steer total demand toward an upper terminal state a_H, while a low value should steer it toward a lower terminal state a_L.

(a) Define the bridge probability

$$h_n^a(y) = \mathbb{P}(Z_N = a \mid Z_n = y).$$

(b) Derive the one-step transition probability of a random walk conditioned to end at a.

(c) Explain how the insider uses trades to make Y behave like this bridge conditional on the value.

(d) Explain why, unconditionally, the market maker can still see total demand with the same law as noise demand.

Solution.

For a simple symmetric random walk, the bridge probability is

$$h_n^a(y) = \mathbb{P}_y(Z_{N-n} = a - y),$$

where the subscript y means the walk starts from y at date n. Explicitly, if $N-n$ and $a-y$ have the same parity and $|a-y| \le N-n$, then

$$h_n^a(y) = 2^{-(N-n)} \binom{N-n}{\frac{N-n+a-y}{2}},$$

and otherwise $h_n^a(y) = 0$.

Conditioning the random walk to end at a gives the bridge transition probability

$$\mathbb{P}(Y_{n+1} = y+1 \mid Y_n = y,\ Y_N = a) = \frac{h_{n+1}^a(y+1)}{h_{n+1}^a(y+1) + h_{n+1}^a(y-1)}.$$

Similarly,

$$\mathbb{P}(Y_{n+1} = y-1 \mid Y_n = y,\ Y_N = a) = \frac{h_{n+1}^a(y-1)}{h_{n+1}^a(y+1) + h_{n+1}^a(y-1)}.$$

The conditional drift of the bridge is therefore

$$\mathbb{E}[\Delta Y_{n+1} \mid Y_n = y,\ Y_N = a] = \frac{h_{n+1}^a(y+1) - h_{n+1}^a(y-1)}{h_{n+1}^a(y+1) + h_{n+1}^a(y-1)}.$$

The insider observes the value. If the value is high, she trades so that total demand Y is pushed toward the high endpoint a_H; if the value is low, she trades so that Y is pushed toward a_L. Her order offsets or reinforces the noise increment in order to make the realized total-demand process have the bridge transition probabilities associated with the correct terminal endpoint.

The subtle point is that the market maker does not observe the endpoint type directly. In an inconspicuous equilibrium, the mixture over value states is chosen so that the unconditional law of Y matches the original noise-trader law. Conditional on the insider's information, total demand has a bridge drift. Unconditional on that information, total demand still looks like a plain symmetric random walk. Information is hidden in the conditional path structure, not in a detectable change in the public distribution of order flow. □

Takeaway. *Inconspicuous trading is camouflage by distribution matching. The insider can condition total order flow on the asset value through a bridge, while the market maker still sees a process whose unconditional law looks like noise.*

References for this problem.

- Lorenz (2024), for binomial Kyle models, inconspicuous equilibria, and bridge constructions.
- Back (1992), for continuous-time Kyle trading and the origin of inconspicuous trading intuition.
- Cetin and Xing (2013), for related bridge structure in a Poisson Kyle model.

3.3 Permanent and Transient Impact

Problem 3.20 (Core: Permanent versus temporary price impact)**.** Let M_t denote the efficient midprice before a trade and let a signed trade q_t move the future midprice according to

$$M_{t+1} = M_t + \lambda q_t + \epsilon_{t+1}.$$

The execution price also includes temporary impact:

$$P_t = M_t + \lambda q_t + \eta q_t,$$

where $\eta > 0$ is a temporary concession paid only at execution.

(a) Compute the expected immediate execution cost relative to M_t.

(b) Compute the expected future midprice change.

(c) Explain which part is permanent and which part is temporary.

(d) Relate this decomposition to Kyle's λ.

Solution.

The signed execution cost relative to the pre-trade midprice is

$$q_t(P_t - M_t) = q_t(\lambda q_t + \eta q_t) = (\lambda + \eta)q_t^2.$$

This is the immediate cost paid by a buy order or, symmetrically, the concession received less favorably by a sell order.

The expected future midprice change is

$$\mathbb{E}[M_{t+1} - M_t \mid q_t] = \lambda q_t,$$

assuming $\mathbb{E}[\epsilon_{t+1} \mid q_t] = 0$. Thus λq_t is the permanent component: it remains in the efficient price after execution. The term ηq_t affects only the execution price and does not persist in M_{t+1}.

In the static Kyle model, λ is the market maker's Bayesian price response to total order flow. In execution models, empirical impact is often split into a permanent component, closer to information revelation, and a temporary component, closer to liquidity demand, immediacy, and order-book resilience. □

Takeaway. *Not every price concession is information. A large part of execution cost can be temporary liquidity pressure even when the long-run midpoint response is small.*

References for this problem.

- Kyle (1985), for permanent price impact as information revelation.
- Almgren and Chriss (2001), for the execution-cost decomposition into permanent and temporary impact.
- Hasbrouck (2007), for empirical measurement of trade price impact.

Problem 3.21 (Advanced: Transient impact and no-dynamic-arbitrage)**.** Suppose signed trades move prices through a transient impact kernel:

$$p_t = p_0 + \sum_{s \le t} G(t-s) q_s,$$

where $G(k)$ is the remaining impact k periods after a unit trade. Consider a two-period round trip: buy Q at $t = 1$ and sell Q at $t = 2$, so

$$q_1 = Q, \qquad q_2 = -Q.$$

(a) Compute the price impact present at the buy.

(b) Compute the impact remaining from the buy when the sell occurs.

(c) Show why $G(1) > G(0)$ would allow profitable price manipulation.

(d) State the economic content of no-dynamic-arbitrage in this setting.

Solution.

The buy at $t = 1$ pays a price that includes contemporaneous impact

$$G(0)Q.$$

The impact cost of the buy is therefore proportional to

$$Q \cdot G(0)Q = G(0)Q^2.$$

At $t = 2$, before the sell's own contemporaneous impact is included, the remaining impact from the first trade is

$$G(1)Q.$$

If the trader sells into a price that has been pushed up by more than the original cost of the buy, the round trip can generate a profit. In the simplest two-trade calculation, the gain from the elevated sale price is proportional to

$$G(1)Q^2,$$

while the initial impact cost is proportional to $G(0)Q^2$. If

$$G(1) > G(0),$$

then buying first and selling later can be profitable even though the net order is zero and no information has arrived.

No-dynamic-arbitrage rules out such mechanical profits from round trips. In linear transient-impact models, this means the decay kernel cannot have shapes that reward a trader for creating and then unwinding his own price impact. More generally, the cost functional over round-trip strategies must be nonnegative. □

Takeaway. *Transient impact is not free to take any convenient statistical shape. Its decay must be economically disciplined, or the model admits price manipulation.*

References for this problem.

- Gatheral (2010), for the no-dynamic-arbitrage restriction on market impact models.
- Bouchaud, Farmer, and Lillo (2009), for market-impact decay and empirical impact regularities.
- Almgren and Chriss (2001), for the execution-model setting in which temporary and permanent impact are separated.

Problem 3.22 (Advanced: Estimating a propagator model of impact)**.** Let $\epsilon_t \in \{-1, +1\}$ denote signed order flow innovations, and suppose returns satisfy the finite-lag propagator model

$$r_t = \sum_{\ell=0}^{L} g_\ell \epsilon_{t-\ell} + \xi_t, \qquad \mathbb{E}[\xi_t \mid \epsilon_t, \epsilon_{t-1}, \ldots] = 0.$$

(a) Write the regression design for estimating $g_0, \ldots, g_L$.

(b) Define cumulative impact through horizon k.

(c) Explain why using raw trade signs instead of innovations can distort the estimated kernel.

(d) Explain why standard errors should allow serial dependence.

Solution.

For each t, define the regressor vector

$$X_t = (\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-L})'.$$

Then

$$r_t = X_t' g + \xi_t, \qquad g = (g_0, g_1, \ldots, g_L)'.$$

The least-squares estimator is

$$\widehat{g} = \left(\sum_t X_t X_t' \right)^{-1} \left(\sum_t X_t r_t \right),$$

provided the lagged-sign design matrix has full rank.

Cumulative impact through horizon $k \leq L$ is

$$G(k) = \sum_{\ell=0}^{k} g_\ell.$$

This quantity measures how much of the initial trade's effect remains after including subsequent return responses up to lag k.

The innovation condition matters because order signs are highly persistent. If raw signs are used, the lag coefficients may partly capture predictable execution splitting rather than the causal response to unexpected order flow. Long-memory order flow can make a slowly decaying impact kernel appear either too persistent or too mean-reverting depending on the conditioning set.

Serial dependence remains even after including lagged signs, because returns, liquidity, and order-flow intensity are clustered intraday. HAC or block standard errors are therefore more appropriate than iid standard errors for testing whether the kernel decays, reverses, or remains permanent. □

Takeaway. *The propagator model turns market impact into an impulse-response problem. The key econometric choice is whether the impulse is raw order flow or the unexpected component of order flow.*

References for this problem.

- Bouchaud, Gefen, Potters, and Wyart (2004), for the propagator view of trade impact and response.
- Bouchaud, Farmer, and Lillo (2009), for the survey treatment of market impact and order-flow persistence.
- Hasbrouck (1991), for trade innovations and ultimate price impact in a VAR framework.
- Newey and West (1987), for dependence-robust covariance estimation.

3.4 Execution and Metaorder Impact

Problem 3.23 (Core: Optimal execution under linear temporary impact)**.** A trader must sell X shares over two periods. Let x be shares sold now and $X - x$ shares sold next period. Temporary impact cost is quadratic:

$$\eta x^2 + \eta(X - x)^2.$$

If the trader waits, the remaining inventory $X - x$ is exposed to price risk. Represent risk cost by

$$\rho\sigma^2(X - x)^2,$$

where $\rho \geq 0$ is risk aversion. The trader chooses x to minimize

$$C(x) = \eta x^2 + \eta(X - x)^2 + \rho\sigma^2(X - x)^2.$$

(a) Solve for the optimal first-period sale x^*.

(b) Show the risk-neutral benchmark.

(c) Explain why risk aversion front-loads execution.

Solution.

Differentiate the objective:

$$C'(x) = 2\eta x - 2\eta(X - x) - 2\rho\sigma^2(X - x).$$

The first-order condition is

$$\eta x = (\eta + \rho\sigma^2)(X - x).$$

Solving,

$$\eta x = (\eta + \rho\sigma^2)X - (\eta + \rho\sigma^2)x,$$

so

$$(2\eta + \rho\sigma^2)x = (\eta + \rho\sigma^2)X.$$

Thus

$$x^* = \frac{\eta + \rho\sigma^2}{2\eta + \rho\sigma^2} X.$$

If $\rho = 0$, then

$$x^* = \frac{X}{2}.$$

The risk-neutral trader splits the order evenly because the quadratic temporary impact cost is minimized by smoothing trades.

If $\rho > 0$, then

$$\frac{\eta + \rho\sigma^2}{2\eta + \rho\sigma^2} > \frac{1}{2}.$$

The trader sells more than half immediately. Risk aversion front-loads execution because waiting leaves inventory exposed to price variance. □

Takeaway. *Optimal execution is a trade-off between walking slowly to reduce impact and trading quickly to reduce inventory risk. The Almgren-Chriss logic is already visible in the two-period case.*

References for this problem.

- Almgren and Chriss (2001), for the canonical optimal-execution model.
- Gatheral (2010), for the connection between execution models and arbitrage-free market impact.
- Hasbrouck (2007), for empirical implementation-shortfall and impact measurement.

Problem 3.24 (Research: Metaorders and square-root impact)**.** Empirical studies of large parent orders often summarize peak impact by the square-root law

$$I(Q) = Y\sigma \left(\frac{Q}{V}\right)^{1/2},$$

where Q is metaorder size, V is daily volume, σ is daily volatility, and Y is a dimensionless constant.

(a) Compare this law with linear Kyle impact $I(Q) = \lambda Q$.

(b) If marginal impact at cumulative quantity q is $Y\sigma(q/V)^{1/2}$, compute total execution cost up to size Q.

(c) Compute average cost per share.

(d) Explain why concavity matters for empirical market impact.

Solution.

Kyle impact is linear in order size:

$$I(Q) = \lambda Q.$$

The square-root law is concave:

$$I(Q) = Y\sigma \left(\frac{Q}{V}\right)^{1/2}.$$

Doubling Q doubles impact in the linear model but increases impact only by $\sqrt{2}$ under the square-root model.

If marginal impact at cumulative quantity q is

$$Y\sigma \left(\frac{q}{V}\right)^{1/2},$$

then total cost is

$$C(Q) = \int_0^Q Y\sigma \left(\frac{q}{V}\right)^{1/2} dq.$$

Therefore

$$C(Q) = \frac{Y\sigma}{V^{1/2}} \int_0^Q q^{1/2}\, dq = \frac{2}{3} Y\sigma \frac{Q^{3/2}}{V^{1/2}}.$$

Average cost per share is

$$\frac{C(Q)}{Q} = \frac{2}{3} Y\sigma \left(\frac{Q}{V}\right)^{1/2}.$$

Concavity matters because it says large trades are costly, but not in direct proportion to size. This is inconsistent with a naive constant-λ impact curve over large metaorders. It suggests that liquidity supply, execution splitting, order-flow persistence, and resilience all shape measured impact. □

Takeaway. *Kyle gives the clean linear benchmark. Metaorder data force students to ask which scale the linear approximation is meant to describe: one auction, one trade, or a large parent order executed over time.*

References for this problem.

- Kyle (1985), for the linear benchmark.
- Bouchaud, Farmer, and Lillo (2009), for empirical market-impact regularities and metaorder evidence.
- Toth et al. (2011), for the square-root impact mechanism based on liquidity near the current price.

Problem 3.25 (Research: Simulation lab for strategic versus mechanical impact)**.** Design a simulation study comparing three data-generating processes:

$$\text{Kyle:} \qquad q_i = \beta v_i + u_i, \quad p_i = \lambda q_i,$$

$$\text{Permanent mechanical impact:} \qquad \Delta p_t = \lambda q_t + \epsilon_t,$$

and

$$\text{Transient propagator:} \qquad r_t = \sum_{\ell=0}^{L} g_\ell q_{t-\ell} + \xi_t.$$

(a) For each model, identify the object called "impact."

(b) State which regression recovers the impact parameter under correct specification.

(c) Give one diagnostic that separates permanent from transient impact.

(d) Give one diagnostic that separates strategic informed trading from mechanical impact.

Solution.

In the Kyle simulation, impact is the equilibrium pricing coefficient

$$\lambda = \frac{\mathrm{Cov}(v, q)}{\mathrm{Var}(q)}.$$

It is a Bayesian inference coefficient: the price response to total order flow because order flow is informative about v.

In the permanent mechanical-impact model, λ is the direct slope in

$$\Delta p_t = \lambda q_t + \epsilon_t.$$

If q_t is exogenous and $\mathbb{E}[\epsilon_t q_t] = 0$, OLS of Δp_t on q_t recovers λ.

In the transient propagator model, impact is the whole lag profile

$$g_0, g_1, \ldots, g_L,$$

or the cumulative response

$$G(k) = \sum_{\ell=0}^{k} g_\ell.$$

The correct regression includes lagged order flow, not only contemporaneous order flow.

A diagnostic for permanent versus transient impact is the long-horizon response after a signed trade. If cumulative response reverses or decays toward zero, impact is transient. If it settles at a nonzero level, part of impact is permanent.

A diagnostic for strategic informed trading is the relation between order flow and future fundamental value or later price revisions. In Kyle, order flow is informative because it contains the insider's signal. In a purely mechanical impact model with exogenous q_t, order flow moves prices by assumption but need not forecast independent value innovations. □

Takeaway. *Simulation teaches the identification lesson cleanly: the same regression slope can mean Bayesian information revelation, exogenous mechanical impact, or a misspecified average of a transient response curve.*

References for this problem.

- Kyle (1985), for strategic informed trading.
- Hasbrouck (1991), for empirical trade innovations and ultimate price impact.
- Bouchaud, Gefen, Potters, and Wyart (2004), for transient propagator impact.
- Gatheral (2010), for no-dynamic-arbitrage discipline in mechanical impact models.

Bibliography

Almgren, Robert, and Neil Chriss (2001). "Optimal Execution of Portfolio Transactions." *Journal of Risk* 3(2), 5–39. [Canonical optimal-execution model with temporary and permanent impact.]

Back, Kerry (1992). "Insider Trading in Continuous Time." *Review of Financial Studies* 5(3), 387–409. [Continuous-time extension of Kyle-style strategic informed trading.]

Back, Kerry, and Hal Pedersen (1998). "Long-Lived Information and Intraday Patterns." *Journal of Financial Markets* 1(3–4), 385–402. [Deterministic time variation in noise trading and intraday Kyle-style liquidity.]

Bagnoli, Mark, S. Viswanathan, and Craig Holden (2001). "On the Existence of Linear Equilibria in Models of Market Making." *Mathematical Finance* 11(1), 1–31. [Distributional restrictions behind linear Kyle equilibria.]

Bouchaud, Jean-Philippe, J. Doyne Farmer, and Fabrizio Lillo (2009). "How Markets Slowly Digest Changes in Supply and Demand." In *Handbook of Financial Markets: Dynamics and Evolution*, Elsevier, 57–160. [Survey source for market impact, order-flow persistence, and metaorders.]

Bouchaud, Jean-Philippe, Yuval Gefen, Marc Potters, and Matthieu Wyart (2004). "Fluctuations and Response in Financial Markets: The Subtle Nature of 'Random' Price Changes." *Quantitative Finance* 4(2), 176–190. [Primary propagator-impact reference.]

Boulatov, Alex, and Dan Bernhardt (2015). "Robustness of Equilibrium in the Kyle Model of Informed Speculation." *Annals of Finance* 11, 297–318. [Robustness of the linear Kyle–Nash equilibrium to small belief errors.]

Chhaibi, Reda, Ibrahim Ekren, and Eunjung Noh (2025). "Solvability of the Gaussian Kyle Model with Imperfect Information and Risk Aversion." arXiv preprint arXiv:2501.16488. [Modern Kyle extension with imperfect information, CARA risk aversion, filtering, and optimal transport.]

Clark, Peter K. (1973). "A Subordinated Stochastic Process Model With Finite Variance for Speculative Prices." *Econometrica* 41(1), 135–155. [Classic subordinated-price process model linking trading activity and return variance.]

Collin-Dufresne, Pierre, and Vyacheslav Fos (2016). "Insider Trading, Stochastic Liquidity, and Equilibrium Prices." *Econometrica* 84(4), 1441–1475. [Continuous-time Kyle model with stochastic liquidity, liquidity timing, stochastic impact, and endogenous volume-volatility.]

Foster, F. Douglas, and S. Viswanathan (1993). "Variations in Trading Volume, Return Volatility, and Trading Costs: Evidence on Recent Price Formation Models." *Journal of Finance* 48(1), 187–211. [Kyle-style model with correlated payoff and noise-trading components.]

Friedrich, Paul, and Josef Teichmann (2020). "Deep Investing in Kyle's Single Period Model."

arXiv preprint arXiv:2006.13889. [Neural-network replication of single-period Kyle equilibrium and computational extensions.]

Fishman, Michael J., and Kathleen M. Hagerty (1995). "The Mandatory Disclosure of Trades and Market Liquidity." *Review of Financial Studies* 8(3), 637–676. [Related disclosure model with manipulation incentives.]

Gallant, A. Ronald, Peter E. Rossi, and George Tauchen (1992). "Stock Prices and Volume." *Review of Financial Studies* 5(2), 199–242. [Empirical evidence on joint price-volume-volatility dynamics.]

Gatheral, Jim (2010). "No-Dynamic-Arbitrage and Market Impact." *Quantitative Finance* 10(7), 749–759. [Primary source for arbitrage-free restrictions on impact decay.]

Goldstein, Itay, Yan Xiong, and Liyan Yang (2025). "Information Sharing in Financial Markets." *Journal of Financial Economics* 163, 103967. [Modern Kyle-style extension on endogenous information sharing and trading-against-error.]

Hasbrouck, Joel (1991). "Measuring the Information Content of Stock Trades." *Journal of Finance* 46(1), 179–207. [Primary empirical source for trade innovations and ultimate price impact.]

Hasbrouck, Joel (2007). *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading.* Oxford University Press. [Main empirical microstructure textbook for price impact.]

Huddart, Steven, John S. Hughes, and Carolyn B. Levine (2001). "Public Disclosure and Dissimulation of Insider Trades." *Econometrica* 69(3), 665–681. [Kyle-style model of post-trade disclosure, dissimulation, and contrarian insider trades.]

Lamoureux, Christopher G., and William D. Lastrapes (1990). "Heteroskedasticity in Stock Return Data: Volume Versus GARCH Effects." *Journal of Finance* 45(1), 221–229. [Evidence that volume helps explain conditional heteroskedasticity.]

Kyle, Albert S. (1985). "Continuous Auctions and Insider Trading." *Econometrica* 53(6), 1315–1335. [Canonical strategic trading and linear price-impact model.]

Lorenz, Christopher (2024). *Equilibria in Kyle Games: Existence and Construction.* Doctoral dissertation, Goethe University Frankfurt. [Discrete Kyle games, finite-action existence, and inconspicuous binomial Kyle equilibria.]

Noldeke, Georg, and Thomas Troger (2006). "A Characterization of the Distributions that Imply Existence of Linear Equilibria in the Kyle-Model." *Annals of Finance* 2, 73–85. [Elliptical distribution characterization for linear Kyle equilibria.]

Newey, Whitney K., and Kenneth D. West (1987). "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3), 703–708. [HAC inference for impact regressions.]

O'Hara, Maureen (1995). *Market Microstructure Theory.* Blackwell. [Theory textbook treatment of strategic trading and market microstructure.]

Toth, Bence, Yves Lemperiere, Cyril Deremble, Joachim de Lataillade, Julien Kockelkoren, and Jean-Philippe Bouchaud (2011). "Anomalous Price Impact and the Critical Nature of Liquidity in Financial Markets." *Physical Review X* 1(2), 021006. [Square-root impact and liquidity-mechanism reference.]