0.1. Understanding Alpha-Mixing Conditions

Formal Definition and Interpretation

Mathematical Setup

Let $\{X_t\}_{t=-\infty}^{\infty}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, P)$. We define:

- $\mathcal{F}_{-\infty}^{t} = \sigma(..., X_{t-1}, X_t)$: the σ-algebra generated by all events up to time t
- $\mathcal{F}_{t+h}^{\infty} = \sigma(X_{t+h}, X_{t+h+1}, ...)$: the σ-algebra generated by all events from time $t+h$ onward

Alpha-Mixing Coefficient

The α-mixing coefficient is defined as:

$$\alpha(h) = \sup_{A \in \mathcal{F}_{-\infty}^{t}, B \in \mathcal{F}_{t+h}^{\infty}} |P(A \cap B) - P(A)P(B)| \tag{1}$$

Interpretation:

- $P(A \cap B)$ is the joint probability of events A and B
- $P(A)P(B)$ is what the joint probability would be if A and B were independent
- $\alpha(h)$ measures the maximum deviation from independence at lag h
- As $h \to \infty$, $\alpha(h) \to 0$ for mixing processes

Necessity of Alpha-Mixing

Statistical Requirements

Alpha-mixing is needed for:
1. **Law of Large Numbers:**

$$\frac{1}{T}\sum_{t=1}^{T} X_t \xrightarrow{p} E[X_t] \tag{2}$$

2. **Central Limit Theorem:**

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} (X_t - E[X_t]) \xrightarrow{d} N(0, \sigma^2) \tag{3}$$

3. **Moment Bounds:**

$$E|\frac{1}{T}\sum_{t=1}^{T} X_t - E[X_t]|^p \leq CT^{-p/2} \tag{4}$$

Understanding the Paper's Mixing Condition

The condition:

$$\sum_{h=1}^{\infty} h^2 \alpha(h)^{\delta/(2+\delta)} < \infty \tag{5}$$

Component Analysis

1. **The Role of h:**
   - h represents the time lag
   - h^2 ensures rapid decay of dependence
   - Larger h means events further apart in time
2. **The Role of $\alpha(h)$:**
   - Measures dependence at lag h
   - Must decay faster than h^{-2} for summability
   - Typical decay: $\alpha(h) \sim h^{-\beta}$ for some $\beta > 2$
3. **The Role of δ:**
   - Controls moment existence
   - Larger δ means stronger moment conditions
   - Typically $\delta = 2$ for financial applications

Intuitive Examples of Mixing

Financial Market Examples

1. **Market Microstructure Effects:**

$$R_t = \phi R_{t-1} + \epsilon_t, \quad |\phi| < 1 \tag{6}$$

- Bid-ask bounce creates short-term dependence
- Effect dies out exponentially: $\alpha(h) \sim |\phi|^h$

2. **Volatility Clustering:**

$$R_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \omega + \alpha R_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{7}$$

- GARCH processes are α-mixing
- Dependence decays geometrically

Verifying Mixing Conditions in Practice

Statistical Tests

1. **Correlation-based Tests:**

$$\hat{\rho}(h) = \frac{\sum_{t=h+1}^{T}(X_t - \bar{X})(X_{t-h} - \bar{X})}{\sum_{t=1}^{T}(X_t - \bar{X})^2} \tag{8}$$

2. **Mixing Coefficient Estimation:**

$$\hat{\alpha}(h) = \sup_{i,j} |\hat{P}(A_i \cap B_j) - \hat{P}(A_i)\hat{P}(B_j)| \tag{9}$$

Practical Approaches

1. **Graphical Analysis:**
   - Plot ACF/PACF
   - Examine decay patterns
   - Check for long-range dependence
2. **Model-based Verification:**
   - Fit ARMA/GARCH models
   - Check residual properties
   - Verify model stability

Connection to Other Time Series Concepts

Related Dependencies

1. **Relationship to Ergodicity:**

$$\alpha\text{-mixing} \implies \text{ergodicity} \tag{10}$$

2. **Comparison with Other Mixing Types:**

- β-mixing (absolute regularity)
- ϕ-mixing (uniform mixing)
- ρ-mixing (maximal correlation)

Hierarchy of Conditions

$$\text{i.i.d.} \implies \phi\text{-mixing} \implies \rho\text{-mixing} \implies \beta\text{-mixing} \implies \alpha\text{-mixing} \tag{11}$$

Stock Return Properties and Mixing

Empirical Evidence

1. **Return Characteristics:**

- Weak serial correlation in returns
- Strong dependence in volatility
- Leverage effects

2. **Market Efficiency Implications:**

$$\alpha(h) \leq Ch^{-\beta}, \quad \beta > 2 \tag{12}$$

- Consistent with weak-form efficiency
- Allows for volatility clustering
- Permits predictability in higher moments

0.2. Understanding Moment Conditions

Overview of Moment Conditions

The moment conditions in our assumption require finite $(4 + \delta)$-th moments for returns, errors, and factors. Let's understand why each condition is necessary and what it buys us in terms of asymptotic theory.

Detailed Analysis of Each Condition

Condition (a): $E|R_{it}|^{4+\delta} < \infty$

This condition on asset returns is needed for several crucial reasons:

1. **Convergence Rates:**

$$\sqrt{T}(\hat{w}_T - w_0) \xrightarrow{d} N(0, V) \tag{13}$$

The fourth moment ensures:

- Existence of the asymptotic variance V
- Validity of the CLT for sample moments
- Uniform convergence of sample covariances

2. **Berry-Esseen Bounds:**

$$\sup_x |P(\sqrt{T}(\hat{w}_T - w_0) \leq x) - \Phi(x)| \leq \frac{C}{\sqrt{T}} \tag{14}$$

The extra δ moment ($E|R_{it}|^{\delta} < \infty$) provides:

- Better convergence rates
- Uniform integrability
- Tighter finite sample bounds

Condition (b): $E|\epsilon_{it}|^{4+\delta} < \infty$

This condition on error terms is crucial for:

1. **Variance Estimation:**

$$\hat{\Sigma}_T - \Sigma = O_p(T^{-1/2}) \tag{15}$$

Where:

- $\hat{\Sigma}_T$ is the sample variance of errors
- Fourth moments ensure consistency of variance estimators
- Extra δ provides uniform convergence

2. **HAC Estimation:**

$$\|\hat{\Omega}_T - \Omega\|_2 = O_p((T/m_T)^{-1/2} + m_T^{-q}) \tag{16}$$

Where:

- $\hat{\Omega}_T$ is the HAC estimator
- Fourth moments ensure kernel estimator convergence
- δ allows for optimal bandwidth selection

Condition (c): $\sup_t E\|F_t\|^{4+\delta} < \infty$

This condition on factors enables:

1. **Factor Structure Analysis:**

$$R_{it} = \beta_i' F_t + \epsilon_{it} \tag{17}$$

Providing:

- Well-defined factor loadings
- Stable estimation procedures
- Valid cross-sectional inference

2. **Uniform Bounds:**

$$\sup_{t,T} E\|\frac{1}{\sqrt{T}}\sum_{s=1}^{t}(F_sF_s' - E[F_sF_s'])\|_2 < \infty \tag{18}$$

Technical Implications

Why $4 + \delta$ Specifically?

1. **Fourth Moments:**
    - Required for CLT with dependent data
    - Needed for convergence of sample covariances
    - Essential for HAC estimation
2. **The Role of δ:**
    - Provides room for Lyapunov condition
    - Ensures uniform integrability
    - Allows for stronger convergence rates

Practical Considerations

Verification in Financial Data

1. **Return Distributions:**

$$\text{Kurtosis} = \frac{E[R_{it}^4]}{(E[R_{it}^2])^2} \tag{19}$$

Typical findings:

- Daily returns: kurtosis $\approx 5 - 10$
- Weekly returns: kurtosis $\approx 4 - 6$
- Monthly returns: kurtosis $\approx 3 - 4$

2. **Factor Properties:**

$$\text{Tail Index} = \lim_{x \to \infty} \frac{\log P(|F_t| > x)}{\log x} \tag{20}$$

Common observations:

- Market factor: tail index $\approx 4 - 5$
- Size factor: tail index $\approx 3 - 4$
- Value factor: tail index $\approx 4 - 5$

Consequences of Violation

If moment conditions fail:

1. **Statistical Issues:**

- Inconsistent variance estimation
- Invalid confidence intervals
- Poor finite sample properties

2. **Econometric Problems:**

- Unstable parameter estimates
- Unreliable hypothesis tests
- Invalid bootstrap procedures

0.3. Understanding Weight Convergence

Basic Concepts of Convergence

What is Convergence?

In our context, convergence means that our estimated weights (w_T^*) get arbitrarily close to the true weights (w^0) as our sample size (T) increases:

$$\|w_T^* - w^0\| \xrightarrow{p} 0 \tag{21}$$

This means:

- For any small error $\epsilon > 0$
- The probability of being more than ϵ away from w^0
- Goes to zero as $T \to \infty$

Why Do We Need Assumptions 1-3?

Assumption 1: Data Generating Process

$$R_{it} = \mu_i(F_t) + \epsilon_{it} \tag{22}$$

This assumption is needed because:

- Ensures returns have a factor structure
- Guarantees existence of synthetic portfolios
- Provides structure for identification

Assumption 2: Mixing Conditions

$$\sum_{h=1}^{\infty} h^2 \alpha(h)^{\delta/(2+\delta)} < \infty \tag{23}$$

This is crucial because:

- Allows for dependent data
- Ensures sample averages converge
- Permits use of uniform LLN

Assumption 3: Moment Conditions

$$E|R_{it}|^{4+\delta} < \infty \tag{24}$$

Required for:

- Existence of limiting distributions
- Uniform convergence of sample moments
- Well-behaved asymptotic theory

Understanding Uniform Convergence

What is Uniform Convergence?

For functions f_n, f on space $\mathcal{W}$:

$$\sup_{w\in\mathcal{W}} |f_n(w) - f(w)| \xrightarrow{p} 0 \tag{25}$$

Key aspects:

- Convergence happens simultaneously for all w
- Rate of convergence is uniform across $\mathcal{W}$
- Stronger than pointwise convergence

Why Do We Need Uniform Convergence?

Critical because:

- Ensures consistency of extremum estimators
- Prevents convergence from failing at the optimum
- Allows interchange of limits and optimization

The Uniform Law of Large Numbers (ULLN)

What is ULLN?

For a sequence of functions $\{g_t(w)\}$:

$$\sup_{w\in\mathcal{W}} |\frac{1}{T}\sum_{t=1}^{T} g_t(w) - E[g_t(w)]| \xrightarrow{p} 0 \tag{26}$$

Why we need it:

- Ensures objective function converges uniformly
- Provides rate of convergence
- Handles dependent data through mixing

The Second Moment Return Matrix

Definition

The second moment return matrix Σ is:

$$\Sigma = E[R_t R_t'] \tag{27}$$

where $R_t = (R_{1t}, ..., R_{Jt})'$

Positive Definiteness

A matrix Σ is positive definite if:

$$x'\Sigma x > 0 \quad \text{for all } x \neq 0 \tag{28}$$

Why it matters:

- Ensures unique solution exists
- Guarantees identification
- Provides stability for estimation

Establishing Identification

What is Identification?

Identification means:

$$w^0 = \arg\min_{w \in \mathcal{W}} Q(w) \quad \text{uniquely} \tag{29}$$

Where:

- $Q(w)$ is the population objective function
- w^0 is the unique minimizer
- No other weights give same synthetic returns

Role of Positive Definiteness

The objective function can be written as:

$$Q(w) = (w - w^0)'\Sigma(w - w^0) \tag{30}$$

Positive definiteness ensures:

- $Q(w) > 0$ for all $w \neq w^0$
- $Q(w^0) = 0$
- Unique minimum at w^0

Why is the Return Matrix Positive Definite?

Economic Arguments

1. **No Arbitrage:**
   - Perfect correlation implies arbitrage
   - Markets eliminate arbitrage
   - Therefore, returns can't be perfectly correlated
2. **Diversification:**
   - Assets have unique risk components
   - Not all risk can be diversified away
   - Implies linear independence of returns

Statistical Verification

We can verify positive definiteness by:

$$\lambda_{min}(\hat{\Sigma}) > 0 \tag{31}$$

Where:

- λ_{min} is the smallest eigenvalue
- $\hat{\Sigma}$ is the sample covariance
- Test statistic follows chi-square distribution

Full Proof Structure

1. **Show Uniform Convergence:**

$$\sup_{w \in \mathcal{W}} |Q_T(w) - Q(w)| \xrightarrow{p} 0 \tag{32}$$

2. **Apply ULLN:**

$$\|Q_T(w) - Q(w)\|_\infty = O_p(T^{-1/2} \log T) \tag{33}$$

3. **Use Identification:**

- Positive definiteness ensures unique minimum
- ULLN ensures sample objective converges
- Therefore, minimizer converges to w^0

4. **Conclude:**

$$\|w_T^* - w^0\| \xrightarrow{p} 0 \tag{34}$$