Instructor: **Stanislav Anatolyev / Yaroslav Korobka** Date: January 6, 2026
Course: **Econometrics of Macro and Financial Data**

1 Stationarity and ergodicity

Definition 1 (Strict stationarity)**.** Series $\{z_t\}_{t=-\infty}^{\infty}$ is said to be strictly stationary if joint distribution of collection $(z_t, z_{t-1}, \dots, z_{t-k})$ does not depend on t for $\forall\, k$.

Definition 2 (Weak stationarity)**.** Series $\{z_t\}_{t=-\infty}^{\infty}$ is said to be weakly stationary if $\mathbb{E}[z_t]$, var $[z_t]$ and cov $[z_t, z_{t-k}]$ for $\forall\, k$ exist and do not depend on t.

Remark 1. Strict stationarity does not imply weak stationarity (e.g. Cauchy).

Definition 3 (Ergodicity)**.** Series $\{z_t\}_{t=-\infty}^{\infty}$ is said to be ergodic if cov $[g(z_t), h(z_{t+k})] \to 0$ as $k \to \infty$ for $\forall\, g$ and h.

Theorem 2 (Invariance to transformations)**.** *If $\{z_t\}_{t=-\infty}^{\infty}$ is stationary and ergodic, then so is $\{f(z_t, z_{t-1}, \dots)\}_{t=-\infty}^{\infty}$ for $\forall$ measurable function f.*

Example 3. We list some examples of the series:

- non-stationary: $y_t = x_t + \delta \cdot \mathbb{1}\{t \geq t_0\}$, $\mathbb{E}[y_t] = \mathbb{E}[x_t]$ for $t < t_0$ and $\mathbb{E}[y_t] = \mathbb{E}[x_t] + \delta$ for $t \geq t_0$
- non-ergodic: $x_t = Z$, where $Z \sim \mathcal{N}(0,1)$, cov $[x_t, x_{t+k}] =$ var $[Z] = 1 \nrightarrow 0$ as $k \to \infty$
- strong white noise (SWN): $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is i.i.d. series, $\mathbb{E}[\varepsilon_t] = 0$, $\sigma^2 =$ var $[\varepsilon_t]$
- weak white noise (WWN): $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is serially uncorrelated, $\mathbb{E}[\varepsilon_t] = 0$, var $[\varepsilon_t] = \sigma^2$, cov $[\varepsilon_t, \varepsilon_{t-j}] = 0$ for $\forall\, j \neq 0$

Example 4. Consider the Bernoulli process $a_t \in \{-1, +1\}$ with $\mathbb{P}\{a_t = +1\} = 1 - \mathbb{P}\{a_t = -1\} = \frac{1}{2}$, and let $\{\theta_t\}_{t=-\infty}^{+\infty}$ be the standard normal white noise independent of $\{a_t\}_{t=-\infty}^{+\infty}$. Show that the process

$$z_t = (a_t - a_{t-1})^2 + \theta_{t+1}^2$$

is strictly stationary and ergodic. Determine its mean and order of serial correlation (you need not derive the whole ACF).

Solution: The process z_t is strictly stationary and ergodic because it is a measurable function of a jointly strictly stationary and ergodic vector process $(a_t, \theta_t)'$. The mean of both a_t and θ_t is zero, a_t is serially independent, and $a_t^2 = 1$ with probability 1. Hence,

$$\begin{aligned} \mathbb{E}[z_t] &= \mathbb{E}[(a_t - a_{t-1})^2] + \mathbb{E}[\theta_{t+1}^2] \\ &= \mathbb{E}[a_t^2] + \mathbb{E}[a_{t-1}^2] - 2\mathbb{E}[(a_t a_{t-1})] + \text{var}\,[(\theta_{t+1})] \\ &= 3. \end{aligned}$$

Because z_t and z_{t+2} are independent, the order of serial correlation cannot exceed 1. The serial correlation in z_t may come only from the a-part. Let us check if it is not zero:

$$\begin{aligned} \text{cov}\,[(z_t, z_{t+1})] &= \text{cov}\left(a_t^2 + a_{t-1}^2 - 2a_t a_{t-1}, a_t^2 + a_{t+1}^2 - 2a_t a_{t+1}\right) \\ &= \text{var}\left(a_t^2\right) = 0. \end{aligned}$$

This, despite the one-period overlap, there is in fact no serial correlation in the process.

2 Lag operator

Definition 4 (Lag operator)**.** Lag operator L is defined as follows:

$$Lx_t = x_{t-1}, \quad LLx_t = x_{t-2}, \quad \dots, \quad L^k x_t = x_{t-k}.$$

Definition 5 (Lag polynomial)**.** Lag polynomial $\Phi(L)$ of order k is defined as

$$\Phi(L) = 1 - \phi_1 L - \dots - \phi_k L^k,$$

so when applied to x_t we have $\Phi(L)x_t = x_t - \phi_1 x_{t-1} - \dots - \phi_k x_{t-k}$.

Theorem 5 (Fundamental theorem of algebra)**.** $\Phi(L)$ *of order k can be factorized as* $\Phi(L) = \Pi_{i=1}^k (1 - \phi_i L)$.

Example 6. Some examples follow:

- $\Phi(0) = 1$
- $\Phi(1) = 1 - \phi_1 - \dots - \phi_k$
- $\Phi(L)\mu = \mu \cdot \Phi(1)$

3 Autocorrelation function (ACF)

Definition 6 (ACF)**.** We define the autocorrelation function as

$$\text{ACF}(j) = \frac{\text{cov}\,[x_t, x_{t+j}]}{\text{var}\,[x_t]}.$$

Remark 7. ACF makes sense only for stationary and ergodic series. Stationarity is used in the denominator, ergodicity in the numerator.

4 Standard linear processes

1. **autoregression of order 1, AR(1):**

$$x_t = \mu + \phi x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{WWN}.$$

   - $\phi = 1$: random walk (with drift $\mu \neq 0$, without drift $\mu = 0$) $\Rightarrow x_t = x_{t-1} + \varepsilon_t$ (non-stationarity, non-ergodic); can write as $x_t = x_0 + \varepsilon_1 + \dots + \varepsilon_t \Rightarrow \text{var}\,[x_t] = \text{var}\,[x_0] + t\sigma^2 = \text{cov}\,[x_t, x_{t+k}]$ (check this); x_t is not measurable so it does not exist as a random variable, $x_t = \sum_{j=0}^{\infty} \varepsilon_{t-j}$.
   - $|\phi| < 1$ (necessary stationarity condition): moments are

$$m_x := \mathbb{E}[x_t] = \mathbb{E}[x_{t-1}]\phi + \mu \Rightarrow m_x = \frac{\mu}{1-\phi}$$

   for the mean,

$$\sigma_x^2 := \text{var}\,[x_t] = \text{var}\,[x_{t-1}]\phi^2 + \sigma^2 \Rightarrow \sigma_x^2 = \frac{\sigma^2}{1-\phi^2}$$

   for the variance, and

$$\gamma_x(1) := \text{cov}\,[x_t, x_{t+1}] = \text{cov}\,[x_t, \mu + \phi x_t + \varepsilon_{t+1}] = \phi\sigma_x^2$$

   covariances with $\gamma_x(j) = \text{cov}\,[x_t, x_{t+j}] = \phi^j \sigma_x^2$. ACF is then ϕ^j.

   We can also write AR(1) using the lag operator as

$$\Phi(L)x_t = \mu + \varepsilon_t, \quad \Phi(L) = 1 - \phi L.$$

   It follows that

$$\begin{aligned} x_t &= \Phi(L)^{-1}\mu + \Phi(L)^{-1}\varepsilon_t \\ &= \Phi(L)^{-1}\mu + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \end{aligned}$$

   because $\Phi(L)^{-1} = \sum_{j=0}^{\infty} \phi^j L^j$ by Taylor.

2. **autoregression of order p, AR(p):**
$$x_t = \mu + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \text{WWN}.$$
Using lag operator we can write
$$\Phi(L)x_t = \mu + \varepsilon_t, \quad \Phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$$
from which follows
$$x_t = \Phi(1)^{-1}\mu + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}.$$
Stationarity condition: roots of $\Phi(L)$ lie outside the unit circle. For example, for AR(1) we have $1 - \phi L = 0 \Rightarrow L = |\frac{1}{\phi}| \Rightarrow |\phi| < 1$.

3. **moving average MA(1):**
$$x_t = \varepsilon_t - \theta\varepsilon_{t-1}, \quad \varepsilon_t \sim \text{WWN}.$$
Always stationary and ergodic process, $\theta \in (-\infty, \infty)$. Moments are
$$\mathbb{E}[x_t] = 0, \quad \text{var}\,[x_t] = (1+\theta)^2\sigma^2, \quad \text{cov}\,[x_t, x_{t+1}] = -\theta\sigma^2,$$
and $\forall\, |k| > 1$ covariances are 0. If $|\theta| > 1 \Rightarrow$ non-invertible representation of MA(1). That is,
$$x_t = \varepsilon_t - \theta\varepsilon_{t-1} \Rightarrow \varepsilon_t = (1-\theta L)^{-1} x_t = \sum_{j=0}^{\infty} \theta^j x_{t-j} \text{ does not converge.}$$
Solution: find an invertible representation (see Hamilton (1994)).

4. **moving average MA(q):**
$$x_t = \Theta(L)\varepsilon_t, \quad \varepsilon_t \sim \text{WWN}, \quad \Theta(L) := 1 - \theta_1 L - \ldots - \theta_q L^q.$$
Always stationary. Invertible if roots of $\Theta(L)$ lie outside the unit circle. ε_t is called *innovation* if MA(q) is invertible.

5. **ARMA(p, q):**
$$\Phi(L)x_t = \mu + \Theta(L)\varepsilon_t, \quad \varepsilon_t \sim \text{WWN}.$$
Stationarity condition: roots of $\Phi(L)$ should be outside the unit circle. Invertibility condition: roots of $\Theta(L)$ should be outside the unit circle. Non-reducability condition: no common roots of $\Phi(L)$ and $\Theta(L)$.

Example 8. Sum of two independent MA(1) processes is MA(1), that is, $\text{MA}(1) + \text{MA}(1) = \text{MA}(1)$ (see Hamilton (1994) for the proof).

Example 9. Sum of two independent AR(1) processes

- with equal coefficients is AR(1). That is, after summing up
$$(1-\pi L)x_t = u_t$$
$$(1-\rho L)w_t = \eta_t$$
we have $(1-\pi L)(x_t + w_t) = u_t + \eta_t$, which is equivalent to $(1-\pi L)y_t = \varepsilon_t$, that is, AR(1) process.
- with different coefficients is ARMA(2, 1). That is, after summing up
$$(1-\pi L)(1-\rho L)x_t = u_t(1-\rho L)$$
$$(1-\pi L)(1-\rho L)w_t = \eta_t(1-\pi L)$$
we have $(1-\pi L)(1-\rho L)(x_t + w_t) = u_t(1-\rho L) + \eta_t(1-\pi L)$. We have two independent MA(1) processes on the right-hand side which is equal to MA(1) due to Example 8. Using the fact that $(1-\pi L)(1-\rho L) = (1-\phi_1 L - \phi_2 L^2)$, we have
$$(1-\phi_1 L - \phi_2 L^2)y_t = \varepsilon_t(1-\gamma L)$$
which is ARMA(2, 1) process.

Example 10. In general, AR(p) + AR(q) = ARMA($p+q$, $\max\{p,q\}$).

5 Wold decomposition

Suppose $\{x_t\}_{t=-\infty}^{\infty}$ is weakly stationary. Then it can be decomposed as

$$x_t = d_t + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i},$$

where d_t is a deterministic part, $\psi_0 = 1$, $\sum_{i=0}^{\infty} \psi_i^2 < \infty$, ε_t is WWN. We call $\varepsilon_t = x_t - \text{Proj}\{x_t | x_{t-1}, \ldots\}$ the Wold innovation; d_t is perfectly predictable from the past, $d_t = \text{Proj}\{d_t | d_{t-1}, \ldots\}$.

Example 11. Some examples of the Wold decomposition:

- white noise: $\eta_t \Rightarrow d_t = 0$, $\psi_0 = 1$, $\psi_j = 0 \; \forall \; j \geq 1$,
- random variable: $x_t = Z$, $Z \sim \mathcal{N}(0, 1) \Rightarrow d_t = Z$, $\varepsilon_t = 0$,
- AR(1) process: $(1 - \phi L)x_t = \mu + \varepsilon_t$, $|\phi| < 1 \Rightarrow x_t = (1 - \phi L)^{-1}(\mu + \varepsilon_t) = (1 - \phi)^{-1}\mu + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$; here, $d_t = (1 - \phi)^{-1}\mu$ and $\psi_j = \phi^j$, $j \geq 0$.

6 Brownian motion

Exercise 12. Let $B(r)$ be the standard Brownian motion on $[0, +\infty)$. Recall the standard Brownian bridge $BB(r) = B(r) - rB(1)$ on $[0, 1]$, and define it on $[0, +\infty)$ by the same formula. Compute the following quantities:

(a) covariance between $B(r)$ and $BB(s)$ for fixed $s > r \geq 1$;

(b) covariance between $B(s)$ and $BB(r)$ for fixed $s > r \geq 1$;

(c) covariance between $B(r)$ and $BB(s)$ for fixed $s \geq 1 > r$;

(d) covariance between $B(s)$ and $BB(r)$ for fixed $s \geq 1 > r$.

Solution. Using the covariance and the Brownian motion properties, we have

(a)

$$\begin{aligned}
\text{cov}\,[B(r), BB(s)] &= \text{cov}\,[B(r), B(s) - sB(1)] \\
&= \text{cov}\,[B(r), B(s)] - s\text{cov}\,[B(r), B(1)] \\
&= r - s \cdot 1 = r - s,
\end{aligned}$$

(b)

$$\begin{aligned}
\text{cov}\,[B(s), BB(r)] &= \text{cov}\,[B(s), B(r) - rB(1)] \\
&= \text{cov}\,[B(s), B(r)] - r\text{cov}\,[B(s), B(1)] \\
&= r - r = 0,
\end{aligned}$$

(c)

$$\begin{aligned}
\text{cov}\,[B(r), BB(s)] &= \text{cov}\,[B(r), B(s) - sB(1)] \\
&= \text{cov}\,[B(r), B(s)] - s\text{cov}\,[B(r), B(1)] \\
&= r - sr = r(1 - s),
\end{aligned}$$

(d)

$$\begin{aligned}
\text{cov}\,[B(s), BB(r)] &= \text{cov}\,[B(s), B(r) - rB(1)] \\
&= \text{cov}\,[B(s), B(r)] - r\text{cov}\,[B(s), B(1)] \\
&= r - r \cdot 1 = 0.
\end{aligned}$$

Exercise 13. Compute the limit of

$$\frac{1}{T} \sum_{t=1}^{T} x_t \varepsilon_t,$$

where x_t is a driftless random walk with weak white noise innovations ε_t.

Solution. We have that $x_t = x_{t-1} + \varepsilon_t$. So using mnemonic rules for the Brownian motion, we have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} x_t\varepsilon_t &= \frac{1}{T}\sum_{t=1}^{T}(x_{t-1}+\varepsilon_t)\varepsilon_t = \frac{1}{T}\sum_{t=1}^{T} x_{t-1}\varepsilon_t + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 \\ &\xrightarrow{d} \sigma^2\int_0^1 B(r)dB(r) + \sigma^2 = \sigma^2\frac{1}{2}(\chi_1^2 - 1) = \frac{\sigma^2}{2}(\chi_1^2+1),\end{aligned}$$

where the first equality on the second line follows from the Ito's lemma. It states that

$$df(B(r)) = f'(B(r))dB(r) + \frac{1}{2}f''(B(r))dr$$

for some function f. Take $f(x) = x^2$, then

$$\begin{aligned} d(B(r))^2 &= 2B(r)dB(r) + dr, \\ \int_0^1 dB(r)^2 &= 2\int_0^1 B(r)dB(r) + \int_0^1 dr, \\ B(1)^2 - B(0) &= 2\int_0^1 B(r)dB(r) + 1, \\ \int_0^1 B(r)dB(r) &= \frac{1}{2}(\chi_1^2 - 1).\end{aligned}$$

Exercise 14. Suppose x_t is a driftless random walk with weak white noise innovations ε_t, but in the regression one uses a mixture of stochastic and deterministic trends instead of either one:

$$x_t = \rho \cdot t \cdot x_{t-1} + \varepsilon_t.$$

Determine the rate of convergence and asymptotic distribution of the OLS estimator of ρ (without too many formalities; uses mnemonic rules).

Solution. We know that

$$\hat{\rho} = \rho + \frac{\sum_{t=1}^{T} t x_{t-1}\varepsilon_t}{\sum_{t=1}^{T} t^2 x_{t-1}^2}.$$

Using the mnemonic rules, we have

$$\frac{1}{T^2}\sum_{t=1}^{T} t x_{t-1}\varepsilon_t \xrightarrow{d} \sigma^2\int_0^1 rB(r)dB(r),$$

and

$$\frac{1}{T^3}\sum_{t=1}^{T} t^2 x_{t-1}^2 \xrightarrow{d} \sigma^2\int_0^1 r^2 B(r)^2 dr.$$

Hence, $r(T)(\hat{\rho}-\rho)$ converges in distribution to

$$\frac{\int_0^1 rB(r)dB(r)}{\int_0^1 r^2B(r)^2dr}$$

with $r(T) = T$.

Exercise 15. Consider the local level model

$$y_t = \mu_t + \epsilon_t,$$

where the local level μ_t follows

$$\mu_t = \mu_{t-1} + \xi_t,$$

and where ϵ_t and ξ_t are independent (temporally and mutually) "structural" shocks.

1. Which AR, MA, ARMA or ARIMA process does y_t follow? Is the series y_t stationary or not?

2. Find the "structural" impulse responses (i.e. impacts of the "structural" shocks) at all horizons. Compare how they differ for the different "structural" shocks. How would you name these two shocks, judging by what you discovered?

Solution.

1. Take differences $\Delta y_t = \Delta \mu_t + \Delta \epsilon_t$ and $\Delta \mu_t = \xi_t$. Then, $\Delta y_t = \epsilon_t - \epsilon_{t-1} + \xi_t$, that is y_t is an ARIMA(0, 1, 1) process. Because y_t contains a unit root, it is not stationary.

2. For ϵ_t, we have $\partial y_t / \partial \epsilon_t = 1$ and $\partial y_{t+k} / \partial \epsilon_t = 0$ for $\forall k \geq 1$, hence we might label it as a temporal shock. For ξ_t, we have $\partial y_t / \partial \xi_t = 1$ and $\partial y_{t+k} / \partial \xi_t = 1$ for $\forall k \geq 1$, hence we might label it as a permanent shock.

Exercise 16. True, False, Unicertain? Explain.

1. "The serial dependence structure of a process is fully characterized by its ACF".

2. "The ACF and IRF cannot exceed 1 in absolute value".

3. "The more linear AR(1) processes are added together, the higher the order of the resulting AR-process will be".

4. "If the DGP is gaussian linear AR, a model cannot be nonlinear".

5. "A strictly stationary series can exhibit explosive behavior in some periods".

6. "If the process is strictly stationary, it is also mean reverting".

7. "If the process has a unit root, the shocks may never die out".

8. "We can use AIC or BIC to choose between models not only in-sample, but also out-of-sample".

9. "The standard Brownian motion is the only Gaussian continuous-time process that equals 0 at $r = 0$ and has variance 1 at $r = 1$".

Exercise 17. Suppose you have specified the following dynamic model for 3-month Treasury bills rates supported by one of financial theories:

$$r_{t+1} - r_t = \varepsilon_{t+1} + \begin{cases} \alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 r_t^{-1}, & \text{if } r_t - r_{t-1} < 0 \\ \beta_0 + \beta_1 r_t + \beta_2 r_t^2 + \beta_3 r_t^{-1}, & \text{if } r_t - r_{t-1} \geq 0 \end{cases}$$

where $\mathbb{E}\left[\varepsilon_{t+1} | r_t, r_{t-1}, ...\right] = 0$ and $\mathbb{E}\left[\varepsilon_{t+1}^2 | r_t, r_{t-1}, ...\right] = \sigma^2 r_t^{2\gamma}$, $\gamma \geq 0$.

Identify all principal differences in how this model differs from the basic two-regime SETAR model that we studied in class. What consequences do these differences have for estimation of the model, inference about its parameters, inference about number of regimes (two vs. one), and forecasting?

Exercise 18. Discuss the questions based on the article "Threshold Autoregressions for Strongly Autocorrelated Time Series" (2002) by M. Lanne and P. Saikkonen.

1. How does a threshold autoregressive (TAR) model differ from traditional linear autoregressive (AR) models, particularly in the context of strongly autocorrelated time series? What are the advantages of TAR models when handling economic time series?

2. Under equation (2), the authors mention that "... time series plots of y_t and z_t can be used as an informal practical aid to assess the significance of the estimated level shifts". Elaborate, how?

3. Discuss the role of threshold parameters in TAR models. What challenges arise in estimating these parameters, and how does the estimation process (e.g., grid search) affect the final model interpretation?

4. In the paper, TAR models are applied to nominal interest rate data. What does the analysis reveal about interest rate behavior? What other macroeconomic quantities can you think of that exhibit similar behavior?

Exercise 19. The following questions follow the article "Credit and Economic Activity: Credit Regimes and Nonlinear Propagation of Shocks" by N. Balke, published in *Review of Economics and Statistics, Vol. LXXXII, N.2, May 2000*.

1. From the description in section II, write out explicitly the structural TVAR.
2. How do you understand the term "propagation of shocks"? What does it mean for the propagation mechanism to be linear or nonlinear?
3. Which of Figures 1-4 can serve as evidence that credit variables are endogenous? Why?
4. Interpret impulse response functions in Figure 2. Is it in line with modern macroeconomic theory?

Exercise 20. The following questions follow the article "The Structure of Unemployment" by D. Papell, J. Murray, and H. Ghiblawi, published in *Review of Economics and Statistics, Vol. LXXXII, N.2, May 2000*.

1. What made the authors not trust the results of unit root ADF tests in (1)? Can you suggest another alternative to proceed than the one the authors chose?
2. Why did the authors choose this sequence of models to consider in this order: (a) no structural breaks, (b) single structural change, (c) multiple structural change?
3. From the evidence on clustering of the break dates, the authors conclude that this is accordant with that "increases in unemployment during recessions have led to increases in long-term unemployment". Explain the logic behind this statement.
4. Why did the authors decide that their results support the structuralist theories of unemployment?
5. What does it mean for an estimate to be median-unbiased? What are the advantages of using α_{MU} and HL_{MU} over using α_{OLS} and HU_{OLS} in section II?
6. Explain why the authors have to construct the $\sup F$ statistic in testing for multiple structural changes (section IV) instead of using the conventional F statistic.
7. When characterizing the unemployment hysteresis, the authors mention the sum of coefficients being close but not exactly equal to 1. What coefficients do they mean? Why do they compare their sum to 1?

Exercise 21. Consider a random walk x_t having innovation ε_t and a stationary variable y_t. Let the variables ε_t and y_t be jointly strictly stationary and correlated at all lags and leads, i.e. $\mathbb{E}[\varepsilon_t y_{t+j}] \neq 0$ for all j. Construct the series $z_t = ax_t + y_t$ for some non-zero a. Suppose one linearly regresses z_t on x_t (without an intercept). Show that the OLS estimator is consistent for a, despite the presence of endogeneity. What will be affected by endogeneity? What is the rate of convergence of the OLS estimator?