ECON 469 Midterm Cheat Sheet
Tiffany Yong Page:1

1 Background

Finite distributed lags

$y = \beta_1 + \sum_{i=0}^q \delta_i x_{t-i} + e_t$ (order q).
Impact multiplier: δ_0, immediate change in y from temp. one-unit increase in x.
Dynamic multiplier: δ_i, $i \ge 0$, gives Δy, i periods after temp. one-unit increase x.
Long run multiplier/propensity: $\theta_0 = \sum_{i=1}^q \delta_i$. This is long run Δy due to perm. one-unit increase in x.
Reparametrise the model s.t. $\delta_0 = \theta_0 = \delta_1 - \delta_2$, and we get $y = \beta_1 + \theta_0 x_t + \delta_1(x_{t-1} - x_t) + \dots + \delta_q(x_{t-q} - x_t)$. Then, obtain $\hat\theta_0$ and $se(\hat\theta_0)$ by regression.

Lagged dependent variables

AR(1): $y_t = \beta_1 + \beta_2 y_{t-1} + u_t$. You lose an observation for each lag.
When $|\beta_2|<1$, equiv. to $y_t = \frac{\beta_1}{1-\beta_2} + e_t + \beta_2 e_{t-1} + \dots = \frac{\beta_1}{1-\beta_2} + \sum_{i=0}^T \beta_2^i e_{t-i}$.
OLS estimator: $\hat\beta = (X^TX)^{-1}X^Ty$.
OLS estimators are not unbiased w/ any assumptions, we need large sample analysis, which is harder w/ corr. across time.
When $y = X\beta + e$ for $t \le T$, Gauss-Markov: $rank(X) = k$, $E[e|X]=0$, $Var(e|X) = \sigma^2 I_T$, then OLS estimator is BLUE.
Assuming $E[e_t|X] = 0\ \forall t$ means y_{t-1} can't be a regressor and y_{t-1} cannot impact x_t.
If y_{t-1} affects x_t, then $Cov(x_t, e_{t-1}) \ne 0$.
If y_{t-1} is a reg, like in AR(1), $Cov(e_t, y_t) = Cov(e_t, \frac{\beta_1}{1-\beta_2} + \sum_{i=0}^T \beta_2^i e_{t-i}) = \sigma^2 \ne 0$.
Assuming $Var(e|X) = \sigma^2 I_T$ means $Var(e_t|X) = \sigma^2\ \forall t$ (homo.), and $Cov(e_t, e_s|X) = 0\ \forall t \ne s$ (no serial corr).
Homoskedasticity: $Var(e_t)$ indep. of X and constant over time.
No serial corr.: reg. errors white noise. y_t, x_t can be serially corr.
$Var(\hat\beta|X) = \sigma^2(X^TX)^{-1}$, and an estimate of σ^2 is $\hat\sigma^2 = \frac{1}{T-k}\sum_{t=1}^T \hat e_t^2$.
$se(\hat\beta_j) = \frac{\hat\sigma}{\sqrt{SST_j(1-R_j^2)}}$ where R_j^2 is the R^2 from reg of x_j on remaining regressors and $SST_j = \sum_{t=1}^T (x_{jt} - \bar x_j)^2$.
Above formulas don't apply w/ hetero. and serial corr.

2 Time series analysis

Cov-stationary: $\{y_t\}$ is cov/weakly stationary if $E(y_t) = \mu$ is indep. of t, and $\gamma_{k,t} = Cov(y_t, y_{t-k}) = E[(y_t-\mu)(y_{t-k}-\mu)] = \gamma_k$.
γ_k is autocov. function, $\rho_k = \frac{\gamma_k}{\gamma_0}$ is autocorr. function.
For a stationary process, $\gamma_k = \gamma_{-k}$
When $E[X]=E[Y]=0$, $Cov(X,Y) = E[XY] - E[X]E[Y]$.
White Noise (WN): $\mu = 0$ and $\gamma_k = 0\ \forall k \ne 0$, then $\{y_t\}$ is a white noise process. If indep. over time, then it is indep. white noise.
Strict stationarity: $\{y_t\}$ is strictly stationary if joint distribution $(y_t, y_{t-1}, \dots, y_{t-k})$ indep. of $t\ \forall k$. Strict stnry implies cov-stnry., but opp. is not true, since higher order moments can dep. on time.
For Gaussian process, completely characterised by μ and cov matrix, so strict and weak stnry are equiv.
Any transformation of strict stnry process is also strictly stnry.
Weak dependence: we need $\gamma_k \to 0$ as $k\to\infty$. For LLN and CLT, $\gamma_k \to 0$ fast enough such that $\sum_{k=0}^\infty |\gamma_k| < \infty$.
Under stnry and weak dep., we can show that $\text{plim}\,\hat\gamma_k = \text{plim}\,\frac{1}{T}\sum_{t=1+k}^T (y_t - \bar y)(y_{t-k} - \bar y) = \gamma_k = Cov(y_t, y_{t-k})$.

MA and AR Models

MA(1): $y_t = \mu + \varepsilon + \theta_1\varepsilon_{t-1}$, where ε_t WN.
We can show that $E(y_t) = \mu$, $\gamma_0 = (1+\theta_1^2)\sigma^2$, $\gamma_1 = \theta_1\sigma^2$, $\gamma_k = 0\ \forall k > 1$. So MA(1) stnry and weakly dep $\forall \mu, \theta$.
ρ_1 attains max (0.5) at $\theta_1 = 1$ and min (-0.5) at $\theta_1 = -1$ by FOC. For all ρ_1 between max and min, there are 2 values of θ_1 that produced it, θ_1 and $1/\theta_1$.
$MA(q)$: $y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q}$, ε_t WN. $\gamma_k = 0$ for $k > q$ and stnry.
Causal: Can express y_t as a fn of ε_t and past ε's. $MA(q)$ causal w/ no restrictions.
$MA(\infty)$: $y_t = \mu + \sum_{j=0}^\infty \psi_j\varepsilon_{t-j}$, ε_t WN and $\{\psi_j\}$ sequence s.t. $\psi_0 = 1$. If $\sum_{j=0}^\infty|\psi_j| < \infty$, y_t cov-stnry and weak dep. For weak dep., $\Rightarrow \sum_k |\gamma_k| < \infty$. For cov-stnry, $E(y_t) = \mu$ and $\gamma_k = \sigma^2\sum_{j=0}^\infty \psi_j\psi_{j+k}$ for $k\ge0$.
AR(1): $y_t = c + \phi_1 y_{t-1} + \varepsilon_t$, ε_t WN. Impulse response fn (IRF) at h: $\frac{\partial y_{t+h}}{\partial\varepsilon_t} = \phi_1^h$.
Stnry condition: $|\phi_1|<1$, since we can write y_t as $MA(\infty)$. When $|\phi_1|\ge1$, no causal stnry soln. When $\phi_1 = 1$ and $c=0$, random walk where $y_t = y_{t-1} + \varepsilon_t = y_0 + \varepsilon_1 + \dots + \varepsilon_t$, not stnry since $Var(y_t) = t\sigma^2$.
When $|\phi_1|<1$, $y_t = \frac{c(1-\phi_1^m)}{1-\phi_1} + \phi_1^m y_{t-m} + \sum_{i=0}^{m-1}\phi_1^i\varepsilon_{t-i}$. As $m\to\infty$, we get $y_t = \frac{c}{1-\phi_1} + \sum_{i=0}^\infty \phi_1^i\varepsilon_{t-i}$, an $MA(\infty)$. $|\phi_1|<\infty$ ensures IRF dies out.
Lag operator: $L^jx_t = x_{t-j}$ and $L^ja = a$ if a const. $\frac{1}{1-\phi_1 L} = \sum_{j=0}^\infty \phi_1^j L^j$.
For AR(1), $(1-\phi_1L)y_t = c + \varepsilon_t$.
Stability condition is $|\phi_1|<1$, which is equivalent to $\phi(z) = 1 - \phi_1 z = 0$ when $|z| = |1/\phi_1| > 1$.
For stable AR(1), IRF at h is ϕ_1^h, $LRM = \frac{1}{1-\phi_1}$, $E(y_t) = \frac{c}{1-\phi_1}$, $\gamma_k = \phi_1^k\frac{\sigma^2}{1-\phi_1^2}\ \forall k\ge0$, and $\rho_k = \phi_1^k$, where $\sigma^2 = Var(\varepsilon_t)$.
Mean deriv: $E(y_t) = c + \phi_1E(y_{t-1}) + E(\varepsilon_t) \Rightarrow (1-\phi_1)E(y_t) = c$.
Yule-Walker eqn: $\gamma_k = \phi_1\gamma_{k-1}$.
So $\sum_{k=1}^\infty|\gamma_k| = \gamma_0\sum_{k=0}^\infty|\phi_1|^k = \gamma_0/(1-|\phi_1|) < \infty$, and process is weak. depndt.
AR(p): $y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t$, or in lag op. terms, $(1-\sum_{i=1}^p\phi_iL^i)y_t = c + \varepsilon_t$.
Stnry cond is p roots of $\phi(z) = 1 - \sum_{i=1}^p\phi_iz^i = 0$ outside unit circle ($|z|>1$).
Under stability, $AR(2)$ can be represented as $MA(\infty)$, where $y_t = \frac{c}{1-\phi_1-\phi_2} + \sum_{i=0}^\infty\psi_i\varepsilon_{t-i}$, so it is causal and stationary. ψ_i's are a function of ϕ_1 and ϕ_2 s.t. $\psi_0 = 1$ and $\sum_{i=0}^\infty|\psi_i|<\infty$.
To get MA coeffs, $\phi(L)y_t = \varepsilon_t$ and so $y_t = \phi(L)^{-1}\varepsilon_t$. So $\psi(L) = \phi(L)^{-1}$. Solve $\phi(L)\psi(L) = 1$ to get $\psi(L)$ terms.
W/ MA representation, $E(y_t) = \frac{c}{1-\phi_1-\phi_2}$.
Yule-Walker: $\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2}$.
Invertible: ε_t depends on $y_t, y_{t-1}, \dots$ (can be written as $AR(\infty)$). AR processes always invertible, MA processes only invertible if MA coeffs restricted.
For $MA(1)$, $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$. If $|\theta|<1$, $y_t = \varepsilon_t + \sum_{i=1}^\infty \theta^i y_{t-i}$, an $AR(\infty)$.
$y_t = \varepsilon_t + \theta\varepsilon_{t-1} = \theta(L)\varepsilon_t$, where $\theta(L) = 1 + \theta L$. $\theta(L)^{-1} = \sum_{i=0}^\infty(-\theta L)^i$, so invert. cond. is $\theta(z) = 1 - \sum_{i=1}^q\theta_iz^i = 0$ root s.t. $|z|>1$.
We get $\theta(L)^{-1}y_t = (1-\theta L + \theta^2L^2 - \dots)y_t = \sum_{j=0}^\infty(-\theta)^jy_{t-j} = \varepsilon_t$. ε_t causal fn of y_t.
AR mdls easily estimated by OLS, so turning MA into AR helps us estimate θ.

Probability

$\text{plim}\,\hat\theta_n = \theta$: $\forall\varepsilon>0$, $\lim P(|\hat\theta_n - \theta| < \varepsilon) = 1$.
$\{X_i\}$ rand. samp where $\mu = E(X_i)$. $\text{plim}(1/n\sum_{i=1}^n X_i) = E(X_i) = \mu$.
Continuous mapping theorem: $\forall g\in C^1$, $\text{plim}\, g(\hat\theta_n) = g(\text{plim}\,\hat\theta_n)$.
$\text{plim}\, 1/n\sum_{i=1}^n(X_i - \bar X)^2 = Var(X_i)$, and $\text{plim}\,1/n\sum_{i=1}^n(X_i-\bar X)(Y_i - \bar Y) = Cov(X_i,Y_i)$. Since LLN holds for stnry and weak dep. data, $\text{plim}\,\hat\gamma_k = \text{plim}\,\frac1T\sum_{t=1+k}^T(y_t-\bar y)(y_{t-k}-\bar y) = \gamma_k$.
MSE-convergence: $\lim E[(y_n - y)^2] = 0 \Rightarrow$ convergence in prob. So $E(\bar y) = \mu$ and $Var(\bar y)\to 0$ implies $\text{plim}\,\hat\gamma_k = \gamma_k$.
Weak exogeneity: $E(x_te_t) = 0\ \forall 1\le t\le T$. Implies $Cov(x_{jt}, e_t) = 0\ \forall j,t$, not $Cov(x_{t+1}, e_t) = 0$.
Consistent: $\text{plim}\,\hat\beta = \beta$. OLS $\hat\beta$ is consistent w/ $Cov(x_t,e_t) = 0$, by weak exog.
Conv. in distribution: Let Z_n, Z rvs, $F_n(x) = P(Z_n\le x)$, $F(x) = P(Z\le x)$. $Z_n \xrightarrow{d} Z$ if $\lim F_n(x) = F(x)\ \forall x$ where $F(x)$ cont.
CLT: $\{Z_i\}_{i=1}^n$ iid rvs, $E(Z_i) = \mu < \infty$, $Var(Z_i) = \sigma^2 < \infty$, then $\frac{\bar Z_n - E(\bar Z_n)}{\sqrt{Var(\bar Z_n)}} \xrightarrow{d} N(0,1)$. Equiv. $\sqrt n\frac{\bar Z_n - \mu}{\sigma}\xrightarrow{d} N(0,1)$.
$Z_n\xrightarrow{P} Z \Rightarrow Z_n\xrightarrow{d} Z$.
Asymp. dist of $\hat\beta_2$: $\sqrt n(\hat\beta_2 - \beta_2) = \frac{1/\sqrt n\sum_{i=1}^n(x_i - \bar x)e_i}{1/n\sum_{i=1}^n(x_i-\bar x)^2}$. $1/n\sum_{i=1}^n(x_i-\bar x)^2\xrightarrow{P} Var(x_i)$, constant. Num: $1/\sqrt n\sum_{i=1}^n(x_i - \bar x)e_i = 1/\sqrt n\sum_{i=1}^n(x_i - \mu + \mu - \bar x)e_i = 1/\sqrt n\sum_{i=1}^n(x_i-\mu) + (\mu - \bar x)1/\sqrt n\sum_{i=1}^n e_i \xrightarrow{P} 1/\sqrt n\sum_{i=1}^n(x_i - \mu)\xrightarrow{d} N(0, Var((x_i-\mu)e_i))$.
$\hat\beta_2\xrightarrow{d} N(0,V)$, where $V = \frac{Var((x_i-\mu)e_i)}{(Var(x_i))^2}$.
If homo, $V = \frac{\sigma^2}{Var(x_i)}$. Converg. rate: $V = Var(\sqrt n\hat\beta_2)$, so $Var(\hat\beta_2) = V/n$. Thus, $\hat\beta_2$ is root-n consistent.
Time series case: $\sqrt T(\hat\beta_2 - \beta_2) = \frac{1/\sqrt T\sum_{t=1}^T(x_t - \bar x)e_t}{1/T\sum_{t=1}^T(x_t-\bar x)^2}$. $1/T\sum_{t=1}^T(x_t-\bar x)^2 \xrightarrow{P} A = (Var(x_t))$, const. $1/\sqrt T\sum_{t=1}^T(x_t - \bar x)e_t\xrightarrow{d} N(0,B)$. So, $Var(\sqrt T\hat\beta_2) = A^{-1}BA^{-1}$. $\hat A = 1/T\sum_{t=1}^T(x_t - \bar x)^2$.
Long run var: $B = Var(1/\sqrt T\sum_{t=1}^T z_t)$, where $z_t = (x_t - \mu)e_t$. For iid data, $B = Var((x_t-\mu)e_t)$, where $\mu = E(x_t)$, can estimate w/ White var est.
Parametric: If we know cov. struct of z_t, we can estimate z_t w/ this.
Nonparametric: use HAC estimator
$\bar z = 1/T\sum_{t=1}^T z_t$, w/ $E(z_t) = 0$. $Var(\bar z) = E((1/T\sum_{t=1}^T z_t)^2) = 1/T^2E(\sum_{t=1}^T\sum_{s=1}^T z_tz_s)$. $\sum_{t=1}^T\sum_{s=1}^T z_tz_s = \sum_{t=1}^T z_t^2 + 2\sum_{k=1}^{T-1}\sum_{t=1+k}^T z_tz_{t-k}$. So, $TVar(\bar z) = Var(\sqrt T\bar z) = 1/T\sum_{t=1}^T E(z_t^2) + 2\sum_{k=1}^{T-1}1/T\sum_{t=1+k}^T E(z_tz_{t-k})$. Under stnrty, $Var(\sqrt T\bar z) = \gamma_0 + 2\sum_{k=1}^{T-1}(1-k/T)\gamma_k$.
So, $B = \lim Var(\sqrt T\bar z) = \gamma_0 + 2\sum_{k=1}^\infty\gamma_k$.
Hypo test: $H_0: \beta_2 = 0$ vs $H_1: \beta_2\ne0$. $\sqrt T(\hat\beta_2 - \beta_2)\xrightarrow{d} N(0, A^{-1}BA^{-1})$. Under H_0, $\sqrt T\hat\beta_2/\sqrt V\xrightarrow{d} N(0,1)$. Replace V with $\hat V$ where $\hat V\xrightarrow{P} V$, then $t = \sqrt T\hat\beta_2/\sqrt{\hat V}\xrightarrow{d} N(0,1)$, since $\sqrt V/\sqrt{\hat V}\xrightarrow{P} 1$.
α is $P(\text{reject } H_0|H_0\text{true})$. Reject if $|t| > 1.96$ for $\alpha = 0.05$.

Variance Estimation

B for $AR(1)$: $z_t = \rho z_{t-1} + u_t$, $u_t\sim WN(0,\sigma^2)$, $|\rho|<1$. $\gamma_0 = \sigma^2/(1-\rho^2)$, $\gamma_k = \rho^k\gamma_0$. $B = \frac{\sigma^2}{1-\rho} + 2\sum_{k=1}^\infty\rho^k\frac{\sigma^2}{1-\rho^2} = \frac{\sigma^2}{(1-\rho)^2}$.
$\hat B_{AR(1)} = \frac{\hat\sigma^2}{(1-\hat\rho)^2}$, parametric estimator.
Nonpara: If $\{z_t\}$ stnry with $\sum_{k=1}^\infty\gamma_k<\infty$ (weak dep), $B = \lim Var(\sqrt T\bar z) = \gamma_0 + 2\sum_{k=1}^\infty\gamma_k$, where $\gamma_k = E(z_tz_{t-k})$ if $E(z_t) = 0$. For $k\le T-1$, $\hat\gamma_k = 1/T\sum_{t=1+k}^T z_tz_{t-k}$.
Naive approach, truncate: $\hat B_{trunc} = \hat\gamma_0 + 2\sum_{k=1}^M\hat\gamma_k$ for $M<T$. M is bandwidth/lag truncation param. If $\gamma_k = 0\ \forall|k|>M$, then consistent. Otherwise, $M\to\infty$ to avoid bias since $Bias(\hat B)\approx 1/M$. But if M grows too fast, $Var(\hat B)\propto M/T$ blows up.
Suppose $M=1$, then $\hat B_{trunc} = \hat\gamma_0 + 2\hat\gamma_1$, and $\hat B_{trunc} < 0$ if $\hat\gamma_1 < -1/2\hat\gamma_0$. One soln is introduce weights $k_T(j)$ s.t. estimator is +ve, leading to HAC kernel estimators.
HAC estimator: $\hat B = \hat\gamma_0 + 2\sum_{j=1}^{T-1}k(\frac{j}{M})\hat\gamma_j$, where $k(x)$ is kernel fn, M is bandwith param, and $\hat\gamma_j = 1/T\sum_{t=1+j}^T(z_t - \bar z)(z_{t-j} - \bar z)$.
Trunc. kernel: $k(x) = \begin{cases}1 \text{ if } |x|\le1\\ 0\text{ if }|x|>1.\end{cases}$
Bartlett (NW) ker $k(x) = \begin{cases}1-|x|, |x|\le1\\0, |x|>1.\end{cases}$
Both ker s.t. when $|j/M|>1$, $k(j/M) = 0$, so M also called truncation lag.
Quadratic spectral ker: $k(x) = \frac{25}{12\pi^2x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right)$.
Quad and Bartlett ker guarantee +ve var.
Choose M to minimise MSE of $\hat B$, so $\min E((\hat B - B)^2) = Bias(\hat B)^2 + Var(\hat B)$.
For NW ker, $Bias(\hat B) = c_1(1/M)$, so $Bias(\hat B)^2 = c_1^2(1/M^2)$. $Var(\hat B) = c_2(M/T)$.
So $MSE(\hat B) = c_1^2(1/M^2) + c_2(M/T)$. W/ FOC, we get $M = d(T^{1/3})$, d const dependent on serial corr struct of z_t.
Can estimate d based on $AR(1)$ mdl for z_t w/ coeff ρ. Rule of thumb for NW ker: $d = 0.75T^{1/3}$ when $\rho = 0.5$.
HAC var estim: $\widehat{Var}(\sqrt T\hat\beta_2) = \hat A^{-1}\hat B\hat A^{-1}$, $\hat A$ is samp. var of x_t, $\hat B = \hat\gamma_0 + 2\sum_{j=1}^{T-1}k(j/M)\hat\gamma_j$, where $\hat\gamma_j = 1/T\sum_{t=1+j}^T(x_t-\bar x)(x_{t-j}-\bar x)\hat e_t\hat e_{t-j}$.
Special case: White var estimator, $\hat B = \hat\gamma_0$

Martingales

No serial corr, but not indep over time.
Martingales: z_t vector and x_t element of z_t. Then, $\{x_t\}$ martingale w.r.t. $\{z_t\}$ if $E(x_t|z_{t-1}, z_{t-2}, \dots) = x_{t-1}\ \forall t$.
Here, info set at $t-1$ is $\{z_{t-1}, z_{t-2}, \dots\}$.
When $z_t = x_t$, x_t martingale w.r.t. past (martingale), so $E(x_t|x_{t-1}, \dots) = x_{t-1}\ \forall t$.
Extended LIE: $\mathcal{G}$, $\mathcal{H}$ info sets, then $E(Y|\mathcal{H}) = E(E(Y|\mathcal{G})|\mathcal{H})$.
If $x_t \in z_t$, then $E(x_t|z_{t-1}, \dots) = x_{t-1} \Rightarrow E(x_t|x_{t-1}, \dots) = x_{t-1}$ by extended LIE.
Random Walk (RW): $z_t = z_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iid(0, \sigma^2)$ and z_0 constant. Then, $z_t = z_0 + \varepsilon_1 + \dots + \varepsilon_t$. So, $E(z_t|z_{t-1}, \dots, z_0) = E(z_{t-1}|z_{t-1}, \dots, z_0) + E(\varepsilon_t|z_{t-1}, \dots, z_0) = z_{t-1}$. Note that $E(\varepsilon_t|z_{t-1}, \dots, z_0) = E(\varepsilon_t|\varepsilon_{t-1}, \dots, z_0) = 0$, since ε_t iid. So $E(z_t) = z_0$, $Var(z_t) = Var(\varepsilon_1 + \dots + \varepsilon_t) = t\sigma^2$.
Martingale difference sequence (m.d.s.): $\{z_t\}$ s.t. $E(z_t) = 0$ is a m.d.s. if $E(z_t|\mathcal{F}^{t-1}) = 0\ \forall t$. Ex: innov of random walk, since $E(z_t|\mathcal{F}^{t-1}) = z_{t-1} \Rightarrow E(z_t - z_{t-1}|\mathcal{F}^{t-1}) = 0$.
M.d.s. have no serial corr, so $Cov(z_t, z_{t-k}) = E(z_t z_{t-k}) = 0\ \forall k$. $E(z_t z_{t-k}) = E(E(z_t z_{t-k}|z_{t-k})) = E(z_{t-k}E(z_t|z_{t-k})) = 0$ since $E(z_t|z_{t-k}) = E(E(z_t|\mathcal{F}^{t-1})|z_{t-k}) = 0$ by LIE.
$z_t \sim iid(0, \sigma^2) \Rightarrow z_t$ m.d.s. $E(z_t|\mathcal{F}^{t-1}) = \int z_t f(z_t|\mathcal{F}^{t-1}) dz_t = \int z_t f(z_t) = E(z_t) = 0$.
z_t m.d.s. $\Rightarrow E(z_t) = 0$. $E(z_t) = E(E(z_t|\mathcal{F}^{t-1})) = E(0) = 0$.
Autoregressive conditional heterskedastic process (ARCH): $\{z_t\}$ is ARCH(1) if $z_t = \sqrt{h_t}\varepsilon_t$, $\varepsilon_t \sim iid(0,1)$, $h_t = \alpha_0 + \alpha_1 z_{t-1}^2$, $\alpha_0, \alpha_1 \geq 0$ so $h_t \geq 0$.
ARCH(1) is a mds, so uncorr, but not independent since $z_t^2 \sim AR(1)$.
To see ARCH(1) is mds, note $\mathcal{F}^{t-1} = \{z_{t-1}, \dots, z_1\} = \{\varepsilon_{t-1}, \dots, \varepsilon_2, z_1\}$. Then, $E(z_t|\mathcal{F}^{t-1}) = E(\sqrt{h_t}\varepsilon_t|\mathcal{F}^{t-1}) = \sqrt{h_t}E(\varepsilon_t|\mathcal{F}^{t-1}) = E(\varepsilon_t) = 0$ since ε_t iid and $E(\varepsilon_t) = 0$.
$z_t = f(z_1, \varepsilon_2, \dots, \varepsilon_t)$ for some f, so z_t is mds $\Rightarrow z_t$ is WN (serially uncorr). $Var(z_t|\mathcal{F}^{t-1}) = E(z_t^2|\mathcal{F}^{t-1}) = E(h_t\varepsilon_t^2|\mathcal{F}^{t-1}) = h_t E(\varepsilon_t^2|\mathcal{F}^{t-1}) = h_t$, since $E(\varepsilon_t^2|\mathcal{F}^{t-1}) = E(\varepsilon_t^2) = 1$. So h_t is conditional var of z_t, time dependent.
We get $z_t^2 = \alpha_0 + \alpha_1 z_{t-1}^2 + \eta_t$, where $\eta_t = z_t^2 - h_t = z_t^2 - E(z_t^2|\mathcal{F}^{t-1})$ is a mds, so WN.
$z_t^2 \sim AR(1)$, so ARCH(1) stnry and weak dep. if $|\alpha_1| < 1$ $(\gamma_k = 0, \forall k \neq 0)$.
Under stnrty, $E(z_t^2) = E(E(z_t^2|\mathcal{F}^{t-1})) = E(h_t) = E(h_{t-1})$. So $E(h_t) = \alpha_0 + \alpha_1 E(h_{t-1})$, and $Var(z_t) = E(z_t^2) = E(h_t) = \frac{\alpha_0}{1-\alpha_1}$.
ARCH(p): $z_t = \mu + \sqrt{h_t}\varepsilon_t$, $h_t = \alpha_0 + \alpha_1 z_{t-1}^2 + \dots + \alpha_p z_{t-p}^2$, so $z_t^2 \sim AR(P)$.
Generalised cond. hetero. mdl of order 1,1 (GARCH(1,1)): Generalise ARCH(1) s.t. $h_t = \alpha_0 + \alpha_1 z_{t-1}^2 + \beta_1 h_{t-1}$, $\alpha_0, \alpha_1, \beta_1 \geq 0$ so h_t non-negative.
$h_t = z_t^2 + (h_t - z_t^2) = z_t^2 - \eta_t$, where $\eta_t = z_t^2 - h_t$ is serially uncorr. innov.
$z_t^2 = \alpha_0 + (\alpha_1 + \beta_1) z_{t-1}^t + \eta_t - \beta_1\eta_{t-1}$, so $z_t^2 \sim ARMA(1,1)$.
z_t indep WN $\Rightarrow z_t$ stnry mds $\Rightarrow z_t$ WN.
(G)ARCH stnry mds, but not indep WN, since z_t^2 are ARMA (dependence exists).
If z_t mds, $E(z_t|\mathcal{F}^{t-1}) = 0$. If z_t WN, only $Cov(z_t, z_{t-k}) = 0$, so z_t can be corr w/ non-linear fns of past values.
Unit root process: $\phi(z) = 1 - z$ has unit root. RW model $(y_t = y_{t-1} + \varepsilon_t, \varepsilon_t \sim iid(0, \sigma^2))$ is a special case of unit root.
Highly persistent time series (shocks have permanent effects), not stnry or weakly dep, as seen from autocorr and impulse response. $Var(y_t) = t\sigma^2$, so $Corr(y_t, y_{t+h}) = \sqrt{\frac{t}{t+h}}$, if $Var(y_0) = 0$. Since corr depends on t, the process is not stationary. Corr. can also be made arbitrarily close to 1 for large t, given h.
$y_{t+h} = y_0 + \varepsilon_1 + \dots + \varepsilon_{t+h}$, so $\frac{\partial y_{t+h}}{\partial \varepsilon_t} = 1\ \forall t, h$.
Temp. unit shock at t creates perm. icnrease in y, so IRF is always 1. For stationary AR(1), $\frac{\partial y_{t+h}}{\partial \varepsilon_t} = \phi_1^h \to 0$ as $h \to \infty$ since $|\phi_1| < 1$.

Forecasting

Optimal MSE forecast of y given x is $E(y|x)$. $\forall g$ fn of x, $MSE(g(x)) = E[(y - g(x))^2] \geq MSE(E(y|x)) = E[(y - E(y|x))^2]$.
$MSE(g(x)) = E[(y - g(x))^2] = E[(y - E(y|x) + E(y|x) - g(x))^2] = E[(y - E(y|x))^2] + 2E[(y - E(y|x))(E(y|x) - g(x))] + E[(E(y|x) - g(x))^2]$.
$MSE(E(y|x)) = E[(y - E(y|x))^2]$, $2E[(y - E(y|x))(E(y|x) - g(x))] = 0$ by LIE cond. on x, $E[(E(y|x) - g(x))^2] \geq 0$.
Optimal MSE forecast of y_{t+h} given $\mathcal{F}^t$ is diff for stnry or non-stnry AR(1) mdls.
Stnry AR(1): $y_t = \phi_1 y_{t-1} + \varepsilon_t$, $|\phi_1| < 1$, ε_t iid $(0, \sigma^2)$ (or mds). Then, $\hat{y}_{t+1|t} = E(y_{t+1}|\mathcal{F}^t) = E(\phi_1 y_t + \varepsilon_{t+1}|\mathcal{F}^t) = \phi_1 E(y_t|\mathcal{F}^t) + E(\varepsilon_{t+1}|\mathcal{F}^t) = \phi_1 y_t$. So one-step ahead forecast $(h = 1)$ is $\hat{y}_{t+1|t} = \phi_1 y_t$.
$h = 2$ forecast: $y_{t+2} = \phi_1 y_{t+1} + \varepsilon_{t+2}$. $\hat{y}_{t+2|t} = E(y_{t+2}|\mathcal{F}^t) = \phi_1 E(y_{t+1}|\mathcal{F}^t) + E(\varepsilon_{t+2}|\mathcal{F}^t) = \phi_1 \hat{y}_{t+1|t} = \phi_1^2 y_t$.
Then, $\forall h \geq 1$, $\hat{y}_{t+h|t} = \phi_1^h y_t$. As $h \to \infty$, $\hat{y}_{t+h|t} \to 0 = E(y_{t+h})$ so y_t becomes irrelevant.
RW: $y_t = y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iid(0, \sigma^2)$, y_0 const.
$E(y_{t+h}|\mathcal{F}^t) = y_t\ \forall h \geq 1$, since $y_{t+h} = y_t + \varepsilon_{t+1} + \dots + \varepsilon_{t+h}$ and $E(\varepsilon_{t+j}|\mathcal{F}^t) = 0\ \forall j \geq 1$.
So as $h \to \infty$, best forecast of y_{t+h} given $\mathcal{F}^t$ is y_t, which is a sign of high persistence.
Forecast error is $y_{t+h} - \hat{y}_{t+h|t} = \varepsilon_{t+1} + \dots + \varepsilon_{t+h} \sim MA(h-1)$.
RW w/ drift: $y_t = \alpha + y_{t-1} + \varepsilon_t$. Then, $y_t = t\alpha + y_0 + \varepsilon_1 + \dots + \varepsilon_t$, and $E(y_t) = \alpha t$ assuming $E(y_0) = 0$ (linear trend).
$\hat{y}_{t+h|t} = E(y_{t+h}|\mathcal{F}^t) = h\alpha + y_t$
AR(1) w/ drift: $y_{t+1} = \alpha + \phi_1 y_t + \varepsilon_{t+1}$.
$\hat{y}_{t+h|t} = \alpha(1 + \phi_1 + \dots + \phi_1^{h-1}) = \phi_1^t y_t$. So as $h \to \infty$, this converges to $E(y_t) = \frac{\alpha}{1-\phi_1}$.