Sometimes we find time series with mixed AR and MA properties (ACF and PACF)

We then can use mixed models: ARMA*(p,q)*

$$Y_t = c + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

MODELING THE SAN DIEGO HOUSE PRICE INDEX

8.1 The Data: Quarterly House Prices in San Diego MSA

Stationary transformation: growth rates

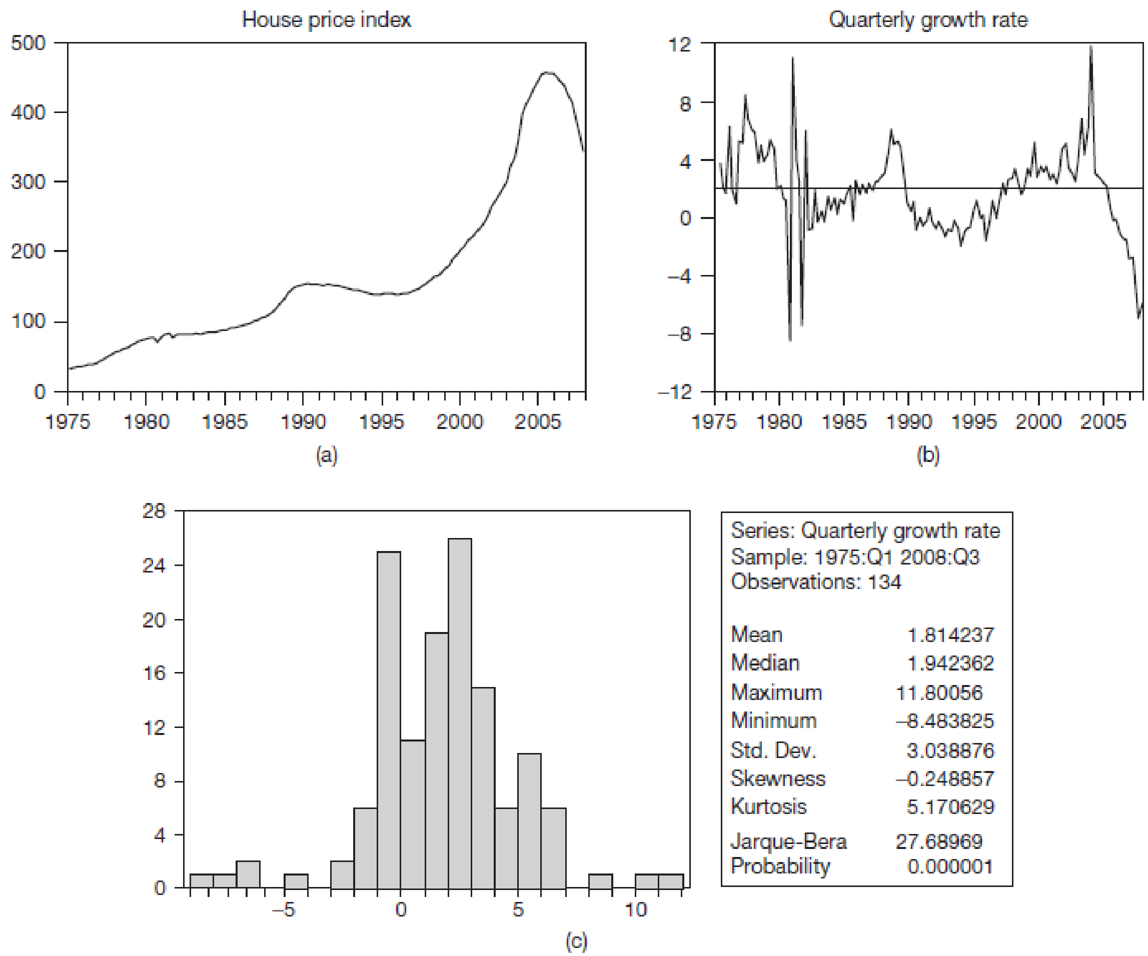
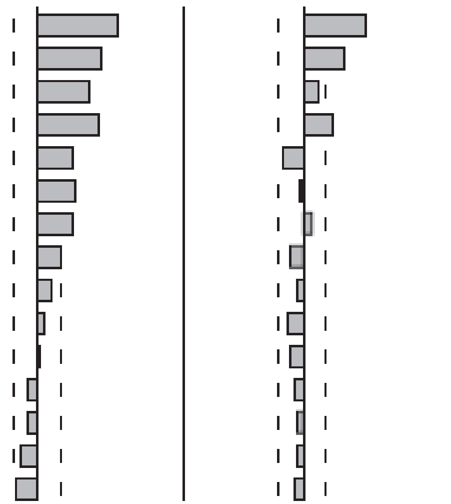


Figure 8.2 Autocorrelation Functions of San Diego Price Growth

Sample: 1975:Q1 2008:Q4
Included observations: 134



	AC	PAC	Q-Stat	Prob
1	0.487	0.487	32.524	0.000
2	0.486	0.326	65.135	0.000
3	0.401	0.121	87.502	0.000
4	0.464	0.223	117.67	0.000
5	0.257	-0.140	127.02	0.000
6	0.276	0.000	137.85	0.000
7	0.264	0.075	147.86	0.000
8	0.184	-0.092	152.77	0.000
9	0.115	-0.040	154.69	0.000
10	0.049	-0.114	155.04	0.000
11	0.011	-0.090	155.06	0.000
12	-0.064	-0.061	155.67	0.000
13	-0.073	-0.025	156.48	0.000
14	-0.123	-0.041	158.77	0.000
15	-0.156	-0.055	162.48	0.000

No clear AR or MA pattern
Certainly significant autocorrelation: Q-stat

$$H_0 : \rho_1 = \rho_2 = \cdots\cdots = \rho_k = 0$$

Why not try a mixed ARMA model?

$$Q_k = T(T+2)\sum_{j=1}^{k}\frac{\hat{\rho}_j^2}{T-j}\,\hat{\rho}_j^2 \rightarrow \chi_k^2$$

8.2 Model Selection

Model 1.	MA(4) $Y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \theta_3\varepsilon_{t-3} + \theta_4\varepsilon_{t-4} + \varepsilon_t$
Model 2.	AR(3) $Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \varepsilon_t$
Model 3.	AR(4) $Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \phi_4 Y_{t-4} + \varepsilon_t$
Model 4.	AR(5) $Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \phi_4 Y_{t-4} + \phi_5 Y_{t-5} + \varepsilon_t$
Model 5.	ARMA(2,2) $Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \varepsilon_t$
Model 6.	ARMA(2,4) $Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \theta_3\varepsilon_{t-3} + \theta_4\varepsilon_{t-4} + \varepsilon_t$

Box/Jenkins Identification/estimation procedure

IDENTIFICATION

ESTIMATION

DIAGNOSTIC CHEKING

USE IT/THEM
or
RESTART

Preliminary transformations Conjecture appropriate models

After estimating the models check:

- Admissibility of each model
  - is it causal/stationary?
  - is it invertible?
- Significance of the parameters
- Whiteness of the residuals
- Explanation power (R^2 or Residual S.E.)

Then, we should compare penalized goodness of fit measures:

AIC, AICc, BIC, SIC, etc.

A process is **covariance stationary (causal)** if it can be written as a linear function of past shocks:

$$X_t = \varepsilon_t + \varphi_1\, \varepsilon_{t-1} + \varphi_2\, \varepsilon_{t-2} + \varphi_3\, \varepsilon_{t-3} + \ldots$$

This happens *iif* all the roots ξ_i of the $\varphi(L)$ polynomial are outside the unit circle:

$$|\xi_i| > 1,$$

i.e., *iff a*ll the modules of the inverse roots are smaller than 1:

$$|1/\xi_i| < 1$$

(if $1/\xi = a + b\,i$, where $i = \sqrt{-1}$, $\sqrt{(a^2 + b^2)} < 1$)

A process is **invertible** if it can be written as a linear function of past observations (up to an unpredictable shock):

$$X_t = \varepsilon_t + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \pi_3 X_{t-3} + \dots$$

This happens *iif* all the roots ξ_i of the $\pi(L)$ polynomial are outside the unit circle:

$$|\xi_i| > 1,$$

I.e., *iff* the modules of the inverse roots are smaller than 1:

$$|1/\xi_i| < 1$$

(if $1/\xi = a + b\,i$, where $i = \sqrt{-1}$, $\sqrt{(a^2 + b^2)} < 1$)

NB: An AR(p) is always *invertible.* A MA(q) is always *stationary.*

8.2.1 Estimation: AR, MA, and ARMA Models

Table 8.1 San Diego House Price Growth, Estimation Output

MA(4)

Dependent Variable: SDG
Method: Least Squares
Sample (adjusted): 1975Q2 2008Q3
Included observations: 134 after adjustments
Convergence achieved after 8 iterations
Backcast: 1974Q2 1975Q1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.684979	0.503973	3.343391	0.0011
MA(1)	0.353809	0.076941	4.598453	0.0000
MA(2)	0.382798	0.082081	4.663663	0.0000
MA(3)	0.234615	0.083348	2.814876	0.0056
MA(4)	0.464352	0.077820	5.966982	0.0000

R-squared	0.394084	Mean dependent var	1.814237
Adjusted R-squared	0.375296	S.D. dependent var	3.038876
S.E. of regression	2.401874	Akaike info criterion	4.626975
Sum squared resid	744.2007	Schwarz criterion	4.735103
Log likelihood	-305.0073	F-statistic	20.97518
Durbin-Watson stat	1.976946	Prob(F-statistic)	0.000000

Inverted MA Roots	.41+.71i	.41-.71i	-.59-.58i	-.59+.58i

AR(4)

Dependent Variable: SDG
Method: Least Squares
Sample (adjusted): 1976Q2 2008Q3
Included observations: 130 after adjustments
Convergence achieved after 6 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.106935	2.627275	0.040702	0.9676
AR(1)	0.238943	0.085584	2.791899	0.0061
AR(2)	0.261707	0.088495	2.957323	0.0037
AR(3)	0.111463	0.089509	1.245274	0.2154
AR(4)	0.284998	0.088905	3.205644	0.0017

R-squared	0.428518	Mean dependent var	1.763910
Adjusted R-squared	0.410231	S.D. dependent var	3.055403
S.E. of regression	2.346440	Akaike info criterion	4.581378
Sum squared resid	688.2228	Schwarz criterion	4.691668
Log likelihood	-292.7896	F-statistic	23.43241
Durbin-Watson stat	1.898907	Prob(F-statistic)	0.000000

Inverted AR Roots	.96	-.01-.65i	-.01+.65i	-.71

Table 8.1 (continued)

ARMA(2,4)

Dependent Variable: SDG
Method: Least Squares
Sample (adjusted): 1975Q4 2008Q3
Included observations: 132 after adjustments
Convergence achieved after 18 iterations
Backcast: 1974Q4 1975Q3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.886929	1.639675	0.540918	0.5895
AR(1)	0.297013	0.196977	1.507853	0.1341
AR(2)	0.580522	0.197836	2.934354	0.0040
MA(1)	-0.007620	0.195334	-0.039012	0.9689
MA(2)	-0.338640	0.151912	-2.229184	0.0276
MA(3)	-0.060909	0.095536	-0.637555	0.5249
MA(4)	0.265362	0.098566	2.692235	0.0081

R-squared	0.431114	Mean dependent var	1.797793
Adjusted R-squared	0.403807	S.D. dependent var	3.057434
S.E. of regression	2.360752	Akaike info criterion	4.607410
Sum squared resid	696.6440	Schwarz criterion	4.760286
Log likelihood	-297.0891	F-statistic	15.78793
Durbin-Watson stat	2.025208	Prob(F-statistic)	0.000000

Inverted AR Roots	.92	-.63		
Inverted MA Roots	.59+.38i	.59-.38i	-.58+.45i	-.58-.45i

8.2.3 Are the Residuals White Noise?

Figure 8.3 San Diego House Price Growth, Correlograms of the Residuals

MA (4)
Sample: 1975:Q2 2008:Q3
Included observations: 134
Q-stastistic probabilities adjusted for 4 ARMA term(s)

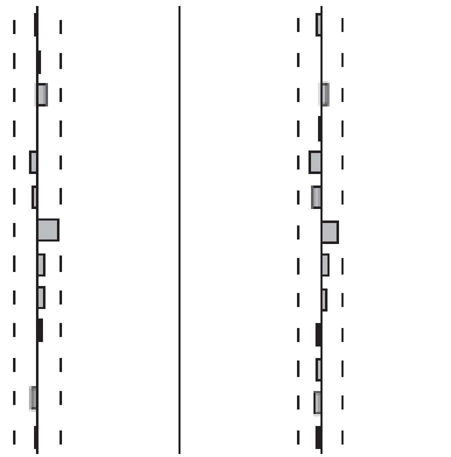
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.004	0.004	0.0022	
		2	0.044	0.044	0.2722	
		3	0.084	0.084	1.2591	
		4	0.046	0.044	1.5514	
		5	0.136	0.131	4.1799	0.041
		6	0.142	0.137	7.0627	0.029
		7	0.125	0.118	9.2956	0.026
		8	0.068	0.048	9.9701	0.041
		9	-0.010	-0.045	9.9835	0.076
		10	-0.009	-0.063	9.9960	0.125
		11	0.024	-0.035	10.081	0.184
		12	-0.031	-0.089	10.224	0.250
		13	-0.016	-0.070	10.262	0.330

AR (4)
Sample: 1976:Q2 2008:Q3
Included observations: 130
Q-stastistic probabilities adjusted for 4 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.043	0.043	0.2433	
		2	0.036	0.034	0.4146	
		3	0.001	-0.002	0.4147	
		4	0.060	0.059	0.9077	
		5	-0.106	-0.111	2.4405	0.118
		6	0.016	0.022	2.4738	0.290
		7	0.156	0.164	5.8619	0.119
		8	0.030	0.010	5.9848	0.200
		9	0.056	0.058	6.4337	0.266
		10	-0.005	-0.025	6.4367	0.376
		11	0.005	-0.016	6.4402	0.489
		12	-0.069	-0.036	7.1398	0.522
		13	-0.025	-0.029	7.2337	0.613

Figure 8.3 (continued)

ARMA (2,4)
Sample: 1975:Q4 2008:Q3
Included observations: 132
Q-stastistic probabilities adjusted for 6 ARMA term(s)



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.019	-0.019	0.0473	
		2	0.014	0.014	0.0742	
		3	0.053	0.054	0.4642	
		4	-0.004	-0.002	0.4666	
		5	-0.061	-0.063	0.9849	
		6	-0.047	-0.052	1.2887	
		7	0.145	0.147	4.2595	0.039
		8	0.048	0.064	4.5857	0.101
		9	0.036	0.038	4.7753	0.189
		10	0.021	-0.002	4.8385	0.304
		11	-0.006	-0.019	4.8438	0.435
		12	-0.048	-0.037	5.1767	0.521
		13	-0.020	-0.001	5.2384	0.631

Criteria to compare models with a different number of parameters

$$AIC = \frac{2m}{T} + \log \frac{\sum_t \hat{\varepsilon}_t^2}{T}$$

$$SIC = \frac{m}{T} \log T + \log \frac{\sum_t \hat{\varepsilon}_t^2}{T}$$

Table 8.2 Summary of Model Estimation and Evaluation

t-statistics

	MA(4)	AR(3)	AR(4)	AR(5)	ARMA(2,2)	ARMA(2,4)
ϕ's (t-ratio) θ's (t-ratio)	$\hat{\theta}_1 = 0.354$ (4.6) $\hat{\theta}_2 = 0.383$ (4.7) $\hat{\theta}_3 = 0.235$ (2.8) $\hat{\theta}_4 = 0.464$ (5.9)	$\hat{\phi}_1 = 0.281$ (3.2) $\hat{\phi}_2 = 0.345$ (3.9) $\hat{\phi}_3 = 0.177$ (1.9)	$\hat{\phi}_1 = 0.238$ (2.8) $\hat{\phi}_2 = 0.261$ (2.9) $\hat{\phi}_3 = 0.111$ (1.2) $\hat{\phi}_4 = 0.284$ (3.2)	$\hat{\phi}_1 = 0.282$ (3.1) $\hat{\phi}_2 = 0.270$ (3.0) $\hat{\phi}_3 = 0.135$ (1.4) $\hat{\phi}_4 = 0.307$ (3.4) $\hat{\phi}_5 = -0.13$ (-1.4)	$\hat{\phi}_1 = 0.134$ (0.6) $\hat{\phi}_2 = 0.777$ (3.5) $\hat{\theta}_1 = 0.137$ (0.5) $\hat{\theta}_2 = -0.415$ (-1.9)	$\hat{\phi}_1 = 0.297$ (1.5) $\hat{\phi}_2 = 0.580$ (2.9) $\hat{\theta}_1 = -0.007$ (-0.04) $\hat{\theta}_2 = -0.338$ (-2.2) $\hat{\theta}_3 = -0.061$ (-0.6) $\hat{\theta}_4 = 0.265$ (2.7)
Covariance-stationary	yes	yes	yes	yes	yes	yes
Invertibility	yes	yes	yes	yes	yes	yes
White noise residuals Q-statistics (p-value)	no $Q_5 = 4.178$ (0.04) $Q_8 = 9.970$ (0.04)	no $Q_4 = 10.722$ (0.0) $Q_8 = 15.099$ (0.01)	yes $Q_5 = 2.440$ (0.1) $Q_8 = 5.984$ (0.2)	yes $Q_6 = 0.704$ (0.4) $Q_8 = 3.968$ (0.2)	no $Q_5 = 7.103$ (0.01) $Q_8 = 11.004$ (0.03)	no $Q_7 = 4.259$ (0.04) $Q_8 = 4.586$ (0.10)
Residual variance $\hat{\sigma}^2_\varepsilon$	5.769	6.003	5.505	5.489	5.784	5.573
Adjusted R-squared	—	0.362	0.410	0.416	—	—
AIC	4.627	4.660	4.581	4.586	4.630	4.607
SIC	4.735	4.747	4.691	4.719	4.739	4.760

$h = 1$	$f_{t,1} = c + \phi_1 Y_t + \phi_2 Y_{t-1} + \phi_3 Y_{t-2} + \phi_4 Y_{t-3}$	$\sigma^2_{t+1\|t} = \sigma^2_\varepsilon$
$h = 2$	$f_{t,2} = c + \phi_1 f_{t,1} + \phi_2 Y_t + \phi_3 Y_{t-1} + \phi_4 Y_{t-2}$	$\sigma^2_{t+2\|t} = \sigma^2_\varepsilon(1 + \phi_1^2)$
$h = 3$	$f_{t,3} = c + \phi_1 f_{t,2} + \phi_2 f_{t,1} + \phi_3 Y_t + \phi_4 Y_{t-1}$	$\sigma^2_{t+3\|t} = \sigma^2_\varepsilon\{1 + \phi_1^2 + (\phi_1^2 + \phi_2)^2\}$

By straight substitution of the estimated parameters of the model and the four most recent values in the information set, we have:

$h = 2008{:}4$	$f_{t,1} = -4.79\%$	$\sigma^2_{t+1\|t} = 2.35^2$
$h = 2009{:}1$	$f_{t,2} = -5.10\%$	$\sigma^2_{t+2\|t} = 2.41^2$
$h = 2009{:}2$	$f_{t,3} = -5.21\%$	$\sigma^2_{t+3\|t} = 2.53^2$

8.3 The Forecast

Figure 8.4 San Diego House Price Growth: Multistep Forecast

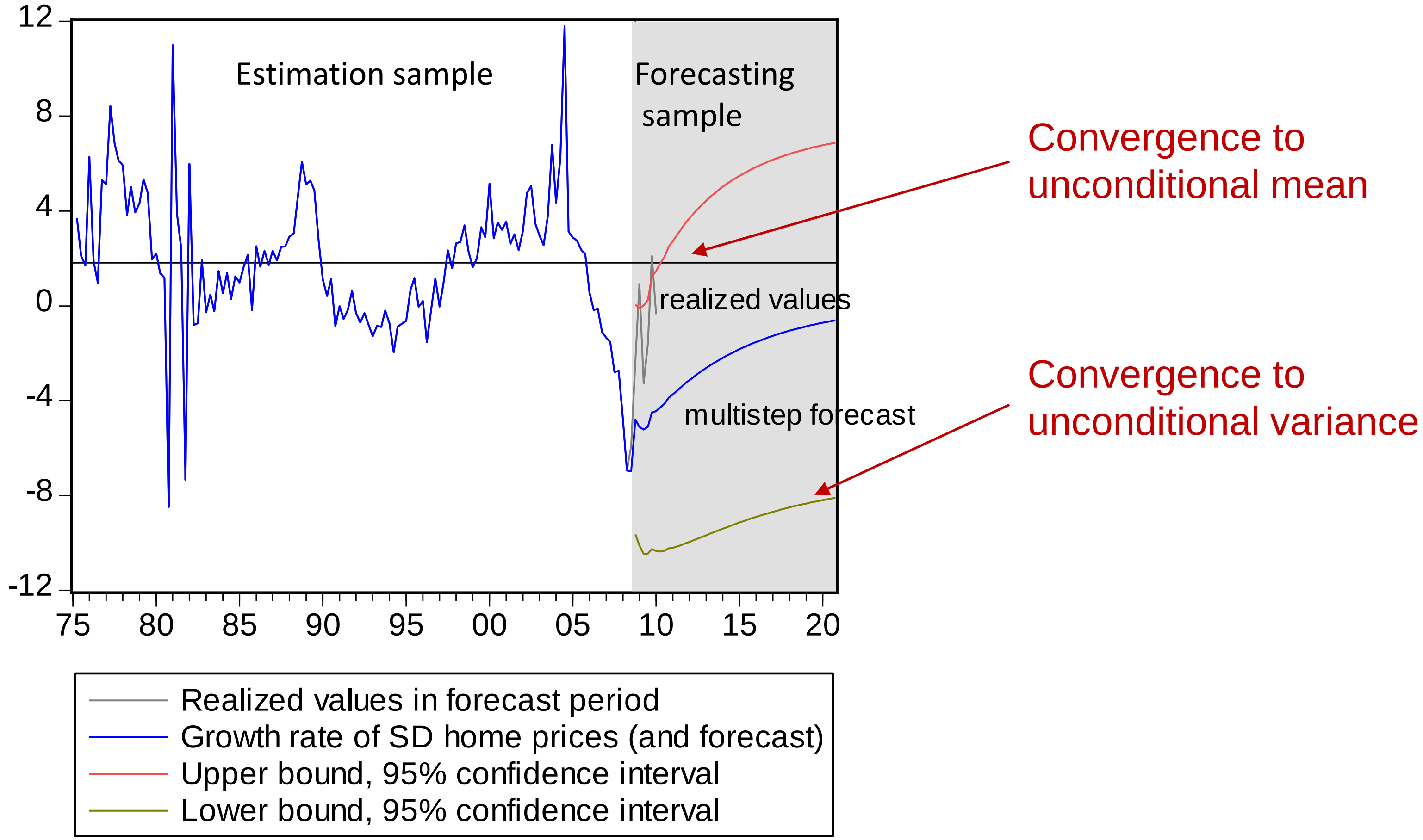


Figure 8.5 Standard Normal Probability Density Function

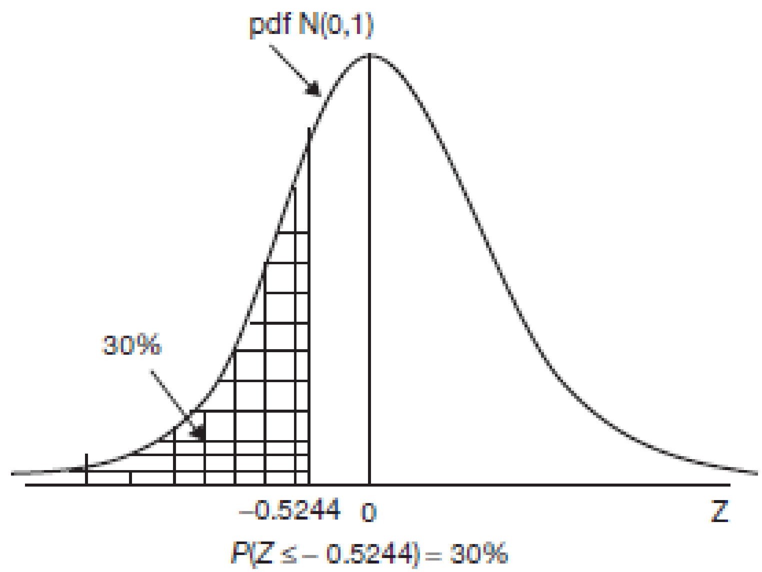


Figure 8.6 February 2009 Density Forecast for San Diego House Price Growth