Ciprian ALEXANDRU,
Ecological University of Bucharest,
Nicoleta CARAGEA,
Ana Maria DOBRE
National Institute of Statistics, Bucharest

## VECTOR AUTOREGRESSIVE MODELS USING "R"

#### **Keywords**

Autoregressive models

Testing

R

vars

JEL Classification C32, O11

#### Abstract

Multivariate data analysis in the context of autoregressive models has evolved as a standard instrument in econometrics. In present, there are developed packages available in R for estimating time series models; one of the most useful package is vars (Pfaff, 2008) containing functions for diagnostic testing, estimation of a restricted models, prediction, causality analysis, impulse response analysis and forecast error variance decomposition. Using the examples provided in the vars vignette, the authors tried to obtain results for the different methods and functions on the base of macroeconomic data set for Romania.

#### 1. INTRODUCTION

In 1980, Sims criticized for the first time the simultaneous equation literature for the ad hoc restrictions needed for identification and for the classification of exogenous and endogenous variables in the system (Baltagi, 2011). In the next period, the multivariate data analysis in the context of vector autoregressive models became a very important tool in applied statistics and econometrics. An argument is the reason that the statistical tests are frequently used in determining inter-dependencies and dynamic relationships between variables, conducting to a fast developing of methodology. Vector Autoregressive models (noted VAR) explain the endogenous variables on the base of their own history, apart from deterministic regressors, as an autoregressive function of their lagged values. Also, the VAR models can include some exogenous variables as like trends and seasonal dummies, but it does not have to classify variables as endogenous or exogenous. The Vector Autoregressive models explain the endogenous variables solely by their own history, apart from deterministic regressors, as

an autoregressive function of their lagged values. Also, the VAR models could include certain exogenous variables like trends and seasonal dummies, but it does not have to classify variables as endogenous or exogenous (Pfaff, 2008).

Thwartly, the structural autoregressive models, SVAR, allow the modelling of contemporaneous explicit interdependence between the left-hand side variables. Hence, these types of models try to bypass the shortcomings of VAR models. At the same time as Sims jeopardized the paradigm of multiple structural equation models laid out by the Cowles Foundation in the 1940s and 1950s, Granger (1981) and Engle Granger (1987) endowed and

With  $A_i$  are (T × T) coefficient matrices for i = 1,...,p and  $u_t$  is a T-dimensional process with  $E(u_t) = 0$  and time invariant positive definite covariance matrix  $E(u_t u_t^K) = \Sigma_u$  (white noise).

One important characteristic of a VAR(p)-process is its stability. This explains

If the solution of the above equation has a root for n = 1, then either some or all variables in the VAR (p) - process are integrated of order one, i.e., I(1). It might be the case of a cointegration between the variables. This instance can then be better analyzed in the context of a vector error correction model (VECM).

With:

Whereby the dimensions of the stacked vectors  $\langle t \rangle$  and  $v_t$  is  $(TP \times 1)$  and the dimension of the matrix A is  $(Tp \times Tp)$ . If the module of the Eigen values of A are less than one, then the VAR(p) – this means the process is stable.

econometricians with a powerful tool for modelling and testing economic relationships, namely, the concept of co-integration.

VAR models have also been used to test the hypothesis that some variables do not have Granger cause to some other variables.

## 2. VECTOR AUTOREGRESSIVE MODELS

A basic VAR model consists of a set of T endogenous variables  $y_k = (y_k, ..., y_{tk}, ..., y_{Tk})$  for t = 1, ..., T. The VAR(p)-process is then defined as:

$$y_k = A_1 y_{k-1} + \dots + A_p y_{k-p} u_t$$
 (1)

that it generates stationary time series with time invariant means, variances and covariance structure, given sufficient starting values. This can be checked by evaluating the characteristic polynomial:

$$\det(I_T - A_1 z - ... - A_n n^p) \neq 0 \text{ for } |n| \leq 1$$
 (2)

In practice, the stability of an empirical VAR (p) - process can be analyzed by considering the companion form and calculating the Eigen values of the coefficient matrix. A VAR(p) - process can be written as a VAR(1) - process:

$$\langle_k = A \langle_{k-1} + v_t \rangle \tag{3}$$

$$\begin{array}{ccc} \mathbf{u_k} \\ \mathbf{v_k} & \mathbf{0} \\ \vdots & & \\ \mathbf{0} & & (4) \end{array}$$

For a given example of the endogenous variables  $y_1,...,y_K$  and sufficient presample values  $y_{-p+1},...,y_0$ , the coefficients of a VAR(p) - process can be estimated efficiently by least-squares applied separately to each of the equations.

Once a VAR(p) model has been estimated, the way is open for further analysis

in near future. A researcher might be interested in diagnostic tests, such as testing for the absence of autocorrelation, heteroscedasticity or non-normality in the error process. Also he might be interested further in causal inference, forecasting and/or diagnosing the empirical  $y_t = \Phi_0 u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + ...$ 

With  $\Phi_0 = I_K$  and  $\Phi_s$  can be computed recursively according to:

For 
$$s = 1, 2, ..., ,$$
  
Whereby  $A_i = 0$  for  $j > p$ .

Finally, the forecasts for horizons  $h \ge 1$  of an empirical VAR(p)-process can be generated recursively according to:

Where  $y_{K+j|K} = y_{K+j}$  for  $j \le 0$ . The forecast error covariance matrix is given as:

and the matrices  $\Phi_i$  are the empirical coefficient matrices of the Wold moving average representation of a stable VAR(p)-process as shown above. The operator  $\otimes$  is the Kronecker product.

VAR models are used for multivariate time series. The structure is that each variable is a linear function of past lags of itself and past lags of the other variables.

Each variable is a linear function of the lag 1 values for all variables in the set.

In a VAR(2) model, the lag 2 values for all variables are added to the right side of the equations. In the case of three s-variables there would be six variables on the right side of each equation, three lag 1 variables and three lag 2 variables.

In general, for a VAR (p) model, the first p lags of each variable in the system would be used as regression predictors for each variable.

VAR models are a specific case of more general vector autoregressive moving-average models (VARMA). A VARMA model for multivariate time series includes the VAR structure above along with moving average terms for each variable. These are special cases

model's dynamic behavior, i.e., impulse response functions and forecast error variance decomposition. The latter two are based upon the Wold moving average decomposition for stable VAR(p)-processes, which is defined as:

$$\Phi_s = \sum_{j=1}^s \Phi_{s-j} A_j \tag{5}$$

$$y_{K+h|K} = A_1 y_{K+h-1/K} + \dots + A_p y_{K+h-p/K}$$
 (6)

As an example, suppose that we measure three different time series variables, denoted by  $s_{t,1}$ ,  $s_{t,2}$ , and  $s_{t,3}$ .

The vector autoregressive model of order 1, denoted as VAR(1), is as follows:

$$s_{t,1} = {}_{1} + {}_{11} s_{t-1,1} + {}_{12} s_{t-1,2} + {}_{13} s_{t-1,3} + w_{t,1} (8) 
s_{t,2} = {}_{2} + {}_{21} s_{t-1,1} + {}_{22} s_{t-1,2} + {}_{23} s_{t-1,3} + w_{t,2} (9) 
s_{t,3} = {}_{3} + {}_{31} s_{t-1,1} + {}_{32} s_{t-1,2} + {}_{33} s_{t-1,3} + w_{t,3} (10)$$

of ARMAX models allowing for the addition of other explicative variables that are outside the multivariate set of principal interest variable.

#### 3. DATA ANALYSIS USING R

A preliminary data analysis is conducted by displaying the summary statistics of the series involved. After that, unit root tests were conducted by applying the Augmented Dickey-Fuller test regressions to the series (ADF test).

We denote the following:

prod = log (GDP/employment);

e = log (employment);

u = unemployment rate;

rw = real wages (log of the real wage index).

The data sets used are from the period 2005-2012, quarterly, for GDP, employment and unemployment rates and real wages (see the *Annex*). The result of the ADF test is presented below, first with trend, then with drift:

Augmented Dickey-Fuller Test Unit Root Test Test regression trend

Call:

 $lm(formula = z.diff \sim z.lag.1 + 1 + tt + z.diff.lag)$ 

#### Residuals:

Min 1Q Median 3Q Max -0.069868 -0.019356 -0.004538 0.014875 0.071735

#### Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.298387 0.929411 1.397 0.176

z.lag.1 -0.133653 0.098952 -1.351 0.190 tt 0.002229 0.002675 0.833 0.413 z.diff.lag1 0.054420 0.203514 0.267 0.792

z.diff.lag2 0.089021 0.201160 0.443 0.662

Residual standard error: 0.03228 on 23 degrees of freedom

Multiple R-squared: 0.1605, Adjusted R-squared: 0.01456

F-statistic: 1.1 on 4 and 23 DF, p-value: 0.3804

Value of test-statistic is: -1.3507 2.7364 1.8605

Critical values for test statistics:

1pct 5pct 10pct tau3 -4.15 -3.50 -3.18 phi2 7.02 5.13 4.31 phi3 9.31 6.73 5.61

> adf2 <- summary(ur.df(diff(dataset[, ''prod'']), type = ''drift'', lags = 1))
> adf2

Augmented Dickey-Fuller Test Unit Root Test Test regression drift

#### Call:

 $lm(formula = z.diff \sim z.lag.1 + 1 + z.diff.lag)$ 

#### Residuals:

Min 1Q Median 3Q Max -0.083811 -0.017178 -0.000067 0.009892 0.092719

#### Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.018768 0.009159 2.049 0.05108 . z.lag.1 -0.795067 0.257340 -3.090 0.00486 \*\* z.diff.lag -0.103986 0.190072 -0.547

0.58917

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03337 on 25 degrees of freedom

Multiple R-squared: 0.4723, Adjusted R-squared: 0.4301

F-statistic: 11.19 on 2 and 25 DF, p-value: 0.0003389

Value of test-statistic is: -3.0896 4.8112

Critical values for test statistics:

1pct 5pct 10pct tau2 -3.58 -2.93 -2.60 phi1 7.06 4.86 3.94

In the next step, an optimal lag length was determined for an unrestricted VAR for a maximal lag length of four.

First, the variables have to be reordered:

> Romania <- dataset[, c("prod", "e", "u", "rw")]

> VARselect(Romania, lag.max = 4, type = "both")

\$selection

 $\begin{array}{ccc} AIC(n) & HQ(n) & SC(n) & FPE(n) \\ 4 & 4 & 1 & 3 \end{array}$ 

#### \$criteria

SC(n) -2.324688e+01 -2.306389e+01 -2.307672e+01 -2.291464e+01 FPE(n) 2.611347e-11 1.636639e-11 1.010456e-11 1.064656e-11

According to the AIC and HQ, the lowest value (Jula, 2010), the optimal lag number is p=4, while SC indicates p=1 and FPE criterion indicates an optimal lag length of p=3.

In the R code lines below is shown the estimation of the VAR(1), as well as the summary output and the diagram of fit for equation "e".

### $p1ct \leftarrow VAR(Romania, p = 1, type = "both")$ p1ct

VAR Estimation Results:

Estimated coefficients for equation prod:

Call:

prod = prod.11 + e.11 + u.11 + rw.11 + const + trend

prod.11 e.11 u.11 rw.11 const trend 0.747003575 0.494499694 -0.011836429 -0.0087230 -5.1660943 0.0053210

Estimated coefficients for equation e:

Call:

e = prod.11 + e.11 + u.11 + rw.11 + const + trend

prod.11 e.11 u.11 rw.11 const trend 0.263865079 0.587760510 0.009050756 -

Estimated coefficients for equation u: Call:

0.1116373 4.674393 -0.00342123

u = prod.11 + e.11 + u.11 + rw.11 + const + trend

prod.11 e.11 u.11 rw.11 const trend -4.56119165 -8.39440840 0.17017893 1.61923859 168.12486 0.09562047

Estimated coefficients for equation rw:

Call:

rw = prod.11 + e.11 + u.11 + rw.11 + const + trend

prod.11 e.11 u.11 rw.11 const trend 0.56619561 0.7138821 0.0049096 0.4906261 -12.9409572 0.0017200

#### > summary(p1ct, equation = "e")

VAR Estimation Results:

Endogenous variables: prod, e, u, rw

Deterministic variables: both

Sample size: 30

Log Likelihood: 207.676

Roots of the characteristic polynomial:

 $0.8216\ 0.7226\ 0.2865\ 0.2865$ 

Call:

VAR(y = Romania, p = 1, type = "both")

Estimation results for equation e:

e=prod.11+e.11+u.11+rw.11+const+trend

| Estimate Std. Error t value $Pr(> t )$ |           |          |        |  |  |  |  |
|--|-----------|----------|--------|--|--|--|--|
| prod.11                                | 0.263865  | 0.105717 | 2.496  |  |  |  |  |
| 0.0198 *                               |           |          |        |  |  |  |  |
| e.11                                   | 0.587761  | 0.174061 | 3.377  |  |  |  |  |
| 0.0025 **                              |           |          |        |  |  |  |  |
| u.11                                   | 0.009051  | 0.006128 | 1.477  |  |  |  |  |
| 0.1527                                 |           |          |        |  |  |  |  |
| rw.11                                  | -0.111637 | 0.071780 | -1.555 |  |  |  |  |
| 0.1330                                 |           |          |        |  |  |  |  |
| const                                  | 4.674393  | 2.824532 | 1.655  |  |  |  |  |
| 0.1110                                 |           |          |        |  |  |  |  |
| trend                                  | -0.003421 | 0.001554 | -2.202 |  |  |  |  |
| 0.0375 *                               |           |          |        |  |  |  |  |
|  |           |          |        |  |  |  |  |

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' '1

Residual standard error: 0.01534 on 24 degrees

of freedom

Multiple R-Squared: 0.5536, Adjusted R-squared: 0.4607

F-statistic: 5.954 on 5 and 24 DF, p-value: 0.001024

Covariance matrix of residuals:

prod u rw prod 0.0009109 -0.0001877 -0.003576 0.0004013 -0.0001877 0.0002354 -0.001708 -0.0001957 -0.0035759 -0.0017082 0.277226 0.0036219 0.0004013 -0.0001957 0.003622 0.0009713

Correlation matrix of residuals:

prod e u rw prod 1.0000 -0.4053 -0.2250 0.4267 e -0.4053 1.0000 -0.2115 -0.4092 u -0.2250 -0.2115 1.0000 0.2207 rw 0.4267 -0.4092 0.2207 1.0000

> *plot(p1ct, names = "e")* 

I vagram of mano residuals for e

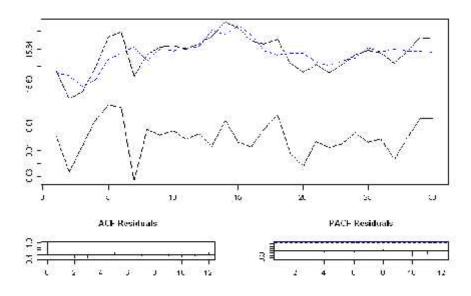


Figure 1. Diagram of fit and residuals for employment

Next, is shown how the diagnostic tests are conducted for the VAR(1) model.

> ser11 <- serial.test(p1ct, lags.pt = 16, type = ''PT.asymptotic'') > ser11\$serial

Portmanteau Test (asymptotic)

data: Residuals of VAR object p1ct Chi-squared = 261.2736, df = 240, p-value = 0.165

> norm1 <- normality.test(p1ct) > norm1\$jb.mul \$JB

JB-Test (multivariate)

data: Residuals of VAR object p1ct Chi-squared = 6.2126, df = 8, p-value = 0.6234

\$Skewness

Skewness only (multivariate)

data: Residuals of VAR object plct

Chi-squared = 3.1601, df = 4, p-value = 0.5314

**\$Kurtosis** 

Kurtosis only (multivariate)

data: Residuals of VAR object p1ct Chi-squared = 3.0525, df = 4, p-value = 0.5491

> summary(ca.jo(Romania, type = ''trace'', ecdet = ''trend'', K = 3,spec = ''transitory''))

Johansen-Procedure

Test type: trace statistic, with linear trend in cointegration

Eigenvalues (lambda):

[1] 8.813106e-01 5.803882e-01 4.306587e-01 2.006071e-01 -2.656685e-16

Values of test statistic and critical values of test:

test 10pct 5pct 1pct r <= 3 | 6.27 10.49 12.25 16.26 r <= 2 | 22.04 22.76 25.32 30.45 r <= 1 | 46.36 39.06 42.44 48.45  $r = 0 \mid 106.03 \ 59.14 \ 62.99 \ 70.05$ 

Eigenvectors, normalised to first column: (These are the cointegration relations)

prod.11 e.11 u.11 rw.11 trend.11 1.0000000 1.00000000 prod.11 1.00000000 1.00000000 1.00000000 e.11 5.6015679 -0.73265559 1.14024691 1.35809064 -0.25451070 0.04961409 u.11 0.7622661  $0.04123177 \ 0.11545465 \ 0.08061866$ 3.0279450 -0.39917112 rw.11 0.61273385 -0.70589985 -0.14974544 trend.11 -0.1098165 -0.01391613 0.01049672 -0.01101037 -0.02518934

Weights W: (This is the loading matrix)

prod.11 u.11 rw.11 e.11 trend.11 prod.d 0.04945092 -0.8389741 0.3683000 0.21454671 -6.022487e-11 0.03468232 2.788080e-11 -0.71099542 -11.7211619 1.3881942 -3.36301607 -3.968509e-10 -0.02179574 -0.6655943 0.6955693 0.14638732 3.625022e-11

# > summary(ca.jo(Romania, type = "trace", ecdet = "trend", K = 2,spec = "transitory"))

Johansen-Procedure

Test type: trace statistic , with linear trend in cointegration

Eigenvalues (lambda):

[1] 7.329818e-01 6.441662e-01 4.021959e-01 1.670903e-01 -2.577691e-17

Values of test statistic and critical values of test:

test 10pct 5pct 1pct  $r <= 3 \mid 5.30 \mid 10.49 \mid 12.25 \mid 16.26$   $r <= 2 \mid 20.22 \mid 22.76 \mid 25.32 \mid 30.45$   $r <= 1 \mid 50.19 \mid 39.06 \mid 42.44 \mid 48.45$   $r = 0 \mid 88.48 \mid 59.14 \mid 62.99 \mid 70.05$ 

Eigenvectors, normalised to first column: (These are the cointegration relations)

e.11 u.11 rw.11 prod.11 trend.11 1.00000000 1.00000000 prod.11 1.000000000 1.00000000 1.00000000 -0.30056061 1.14171040 e.11 1.536746619 0.68044005 7.27785201 u.11 0.17068541 0.04906183 0.001081955 -0.01590393 0.01002381 0.17634015 -0.62296427 rw.11 0.578505618 -0.27935443 -3.96974515 -0.03052553 -0.01081926 trend.11 0.008170070 -0.01126230 0.10388828

Weights W: (This is the loading matrix)

prod.11 e.11 u.11 rw.11 trend.11 prod.d -0.07902084 0.1151782 0.31305758 -0.134844731 -3.295854e-13 e.d 0.04441833 -0.1047434 0.32808259 -0.004793649 4.902170e-13 -5.53342117 -9.0910660 -1.99112917 1.179708859 1.891516e-11 -0.27297765 0.7167966 -0.08963894 -0.051798214 -2.220969e-12

#### 4. CONCLUSIONS

In this paper we emphasize the utility of the functions and methods contained in the package *vars* within R - software environment for statistical computing. By this package, the researchers have a fairly ease to use environment for conducting VAR, SVAR and SVEC analysis. For empirical experimentation we have a determination of a suitable lag length for the model, stability and causality analysis of Romanian macroeconomic data sets.

#### 5. Acknowledgments

The authors would like to thank to Bernhard Pfaff who implemented the *vars* package and also to other reviewers for valuable feedback on this article.

#### References

- [1] Holland F and Fullman D (2010) Local-government data sharing: Evaluating the performance of collaborative innovation *Cross-Cultural Management Journal* **20(7)** pp743-768
- [2] Day RA (1995). *How to write and publish a scientific paper*. Cambridge University Press,

- Cambridge.
- [3] Baltagi, H.B. (2011) *Econometrics*, Springer Publisher, ISBN 978-3-642-20058-8
- [4] Caragea N., Alexandru, C.A., Dobre A.M. (2012) Bringing New Opportunities to Develop Statistical Software and Data Analysis Tools in Romania. In: "Alexandru Ioan Cuza" University of Iasi, Faculty of Economics and Business Administration, 6th edition of the Annual International Conference "Globalization and Higher Education in Economics and Business Administration" -GEBA 2012, Iasi, Romania, 18-20 October 978-973-703-766-4, 2012. **ISBN** Conference Proceedings Citation Index
- [5] Doornik, J. A. and D. F. Hendry (1997) Modelling Dynamic Systems Using PcFiml 9.0 for Windows, International Thomson Business Press, London.
- [6] Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica: 987-1007.
- [7] Engle RF, Granger CWJ (1987) "Co-Integration and Error Correction: Representation, Estimation and Testing." Econometrica, 55, 251-276.
- [8] Granger CWJ (1981) "Some Properties of Time Series Data and Their Use in Econometric Model Specification." Journal of Econometrics, 16, 121-130.
- [9] Hamilton, J. (1994) Time Series Analysis, Princeton University Press, Princeton.
- [10] Jula, N. and Jula, D. (2009) Modelare economica. Modele econometrice si de optimizare., Mustang Publisher, ISBN 978-606-8058-14-6.
- [11] Lütkepohl, H. (2006) New Introduction to Multiple Time Series Analysis, Springer, New York.

- [12] Pfaff, B., (2008) VAR, SVAR and SVEC Models: Implementation Within R Package vars, Journal of Statistical Software
- [13] Sims CA (1980) "Macroeconomics and Reality." *Econometrica*, 48, 1-48

#### **Biography**

Ciprian Alexandru participated in various conferences. His research activity was valued by publishing studies in journals in Romania and Europe, as well as in international databases. He attended a postdoctoral program that implemented the use of R as a tool for analyzing the evolution indices on the stock market.

Nicoleta Caragea participated as a national expert in various projects, workshops and conferences organized by prestigious international institutions with statistics activity. The results of her research activity have been published in many prestigious journals in Romania and abroad and registered in international databases.

Ana Maria Dobre is expert at the National Institute of Statistics within General Department of Social **Statistics** and Demography. Currently she participates in the development of studies on international migration within which was implemented the use of R. Her research activity consists in participation at conferences and publishing working papers

#### Annex

| Y       | Q    | e       | prod   | rw       | u     |  |
|---------|------|---------|--------|----------|-------|--|
| 1 2005  | Q1   | 15.5882 | 9.3495 | 6.8189   | 8.5   |  |
| 2 2005  | Q2   | 15.6141 | 9.3592 | 6.8596   | 7.1   |  |
|         |      | 15.5829 |        |          |       |  |
| 4 2005  | Q4   | 15.5904 | 9.4618 | 6.9441   | 6.8   |  |
| 5 2006  | Q1   | 15.6167 | 9.4879 | 6.9782   | 7.8   |  |
| 6 2006  | Q2   | 15.6538 | 9.4910 | 7.0157   | 7.0   |  |
|         |      | 15.6597 |        |          |       |  |
|         | _    | 15.6078 |        |          |       |  |
|         | _    | 15.6333 |        |          |       |  |
|         | _    | 15.6417 |        |          |       |  |
|         |      | 15.6437 |        |          |       |  |
|         | _    | 15.6397 |        |          |       |  |
|         | _    | 15.6455 |        |          |       |  |
|         | _    | 15.6546 |        |          |       |  |
|         | _    | 15.6709 |        |          |       |  |
|         | _    | 15.6637 |        |          |       |  |
|         | _    | 15.6492 |        |          |       |  |
|         |      | 15.6453 |        |          |       |  |
|         | _    | 15.6505 |        |          |       |  |
|         | _    | 15.6234 |        |          |       |  |
|         | _    | 15.6131 |        |          |       |  |
|         | _    | 15.6219 |        |          |       |  |
|         | _    | 15.6128 |        |          |       |  |
|         | _    | 15.6221 |        |          |       |  |
|         |      | 15.6332 |        |          |       |  |
|         |      | 15.6380 |        |          |       |  |
|         |      | 15.6355 |        |          |       |  |
|         |      | 15.6233 |        |          |       |  |
|         | _    | 15.6378 |        |          |       |  |
|         | _    | 15.6530 |        |          |       |  |
| 31 2012 | 2 Q3 | 15.6513 | 10.107 | 4 7.663. | 3 6.8 |  |