VAR PLS

Graciela González Farías

Francisco Corona and Jesús Gonzalo*
Centro de Investigación en Matemáticas, Campus Monterrey
*Universidad Carlos III de Madrid

ISBIS 2012

Bangkok, Thailand, June 17, 2012

Outline

- Motivation
- PLSAR and PLS
- Quick overview of VAR Models
- Definition for VAR-PLS
- Bootstrap for VAR-PLS
- Conclusions

Motivation

- PLS is a technique that has had a big impact on many applications such as quality control for batch processes in Chemical industries, medical image analysis, microarrays, path modeling, classification, discrimination, spatio-temporal PLS models, just to mention some, with authors such as McGregor, Nomikos, MacIntosh, V. Esposito Vinz, P. Garthwaite, among others
- It can be used in univariate and multivariate data as well
- *It has been shown that gives better predictions even when the standard assumptions are met*
- Phillip Hans Franses (2006) proposes a methodology to construct the forecast h steps ahead in an optimal way, through an autoregressive order p model: An *Autoregressive Partial Least Squares Model denoted as* $PLSAR(h, p)$

Case Study

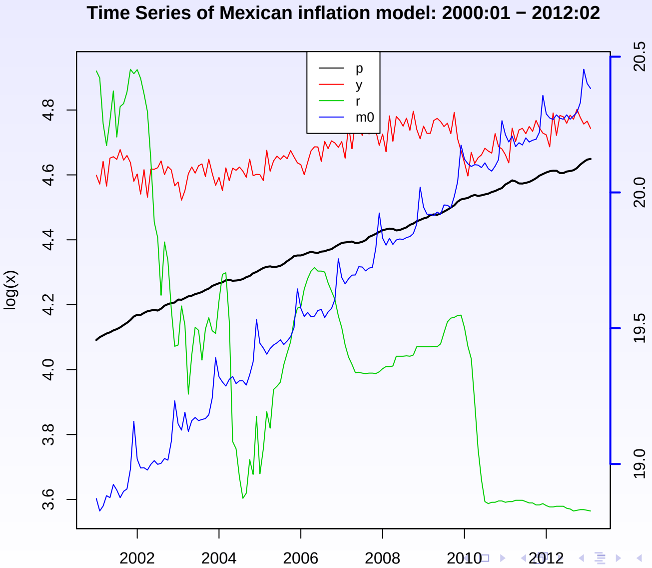
- To develop a model to predict the Mexican inflation, as precise as possible.
- The model has to consider, as the principal source of the Mexican inflation, the growth and variation of the monetary condition of the country.
- Irrespective of all possible discussions, there seems to be a common understanding that the inflationary process, in the lung run, is a purely monetary phenomena.
- Here, we will not discuss whether or not such relationship exists but we will show its empirical properties with a model that is tested out of sample via its error prediction measure.

Case Study

We work with 4 indexes (we built them) from January 2000 to Feb 2012

- p: Consumer price index.
- $m0$: Monetary base.
- r: Equilibrium interest rate (28 days).
- y: Industrial production index.

Case Study



Time Series of Mexican inflation model: 2000:01 − 2012:02
p
y
r
m0
log(x)
3.6
3.8
4.0
4.2
4.4
4.6
4.8
19.0
19.5
20.0
20.5
2002
2004
2006
2008
2010
2012

Objective

- We generalize the work proposed by Franses in the following way:

  1. Give a multivariate representation based on the flexibility of VAR models, model that we will call $VAR - PLS(h, p)$.
  2. Extend the model to consider deterministic variables (dummies, trend, etc.) as well as exogenous variables.
  3. Bootstrap prediction intervals.
  4. Compare the forecast capabilities between $VAR - PLS(h, p)$ models and $VAR(p)$ models explicitly built for prediction (integral predictor method).

$$PLSAR(h, p)$$

Franses considers three different ways to construct a forecast for an $AR(p)$:

1- A single model for all horizons

$$AR(p) : y_{T+h} = \mu + \rho_1 y_{T+h-1} + \rho_2 y_{T+h-2} + \cdots + \rho_p y_{T+h-p} + \epsilon_T$$

For the $AR(p)$ this is the classical iterative procedure to get h steps ahead forecast, plus the fact that we estimate the parameters via LS.

2- One model for each horizon, the variance will vary within each horizon and *there are different models for each step*

$$AR_h(p) : y_{t+h} = \mu + \rho_{1,h} y_t + \rho_{2,h} y_{t-1} + \cdots + \rho_p y_{t-p} + \epsilon_{t,h}$$

The $AR_h(p)$ is an alternative to the $AR(p)$ because

- LS minimizes the sum of squares of ϵ_t's but there is no way to assure it will remain minimum for all the h steps in the future.
- For stationary time series recall the forecast of an $AR(p)$ model quickly converges to the inconditional mean (and variance, for the interval prediction error) clearly depending on $h \geq p$.
- For more details on this type of models see : Pesaran & Pick (2010), Marcellino, Stock & Watson (2004), Carreiro, Kapetorios & Marcellino (2010), Tiao & Xu (1993) among others.

3- Something in between: *PLSAR*, this model behaves like *in the middle* between an $AR(p)$ and a $AR_h(p)$

$$PLSAR(h, p) : \hat{Y} = XB_{PLS}$$

- It is clear that there exists adjacent correlation between the time series, and neither one of the above models takes it into account. In other words, we know that (y_t, y_{t-j}) are correlated and so are (y_{T+h}, y_{T+h-j}). Therefore we would like to jointly predict $(y_{t+h}, y_{t+h-1}, y_{t+h-2}, \ldots, y_{t+1})$ through $(y_t, y_{t-1}, y_{t-2}, \ldots, y_{t-p})$. PLS is a technique very appealing to do so.

$PLSAR(h, p)$

3- Franses proposed to arrange the information as

$$(y_t, y_{t-1}, y_{t-2}, \ldots, y_{t-p}) \text{ as the predictor matrix } X$$

$$(y_{t+h}, y_{t+h-1}, y_{t+h-2}, \ldots, y_{t+1}) \text{ as the predicted matrix } Y$$

- Apply the *PLS* algorithm to get the latent variables with the relevant information given in X and Y.
- His simulations show that the $PLSAR(h, p)$ is quite competitive with respect to the classical models in the literature.

PLS

PLS

- PLS can be seen from different viewpoints. For us, the relationship between its linear expression will be the best one, in order to relate it with a Vector Autoregressive model

$$Y = XB + U,$$

where Y is a $N \times k$ matrix, X is $N \times p$, B is a $p \times k$ matrix, and U is $N \times k$.

PLS

- The basic procedure maximizes a covariance

$$\max cov(X\alpha, Y\beta)^2$$

under certain restrictions,

$$\alpha'(S^*_{xx} + \lambda_x)\alpha = 1 \quad \text{and} \quad \beta'(S^*_{yy} + \lambda_y)\beta = 1$$

where $S^*_{xx} = (1 - \lambda_x)S_{xx}$ and $S^*_{yy} = (1 - \lambda_y)S_{yy}$.

  - $(X\alpha, Y\beta)$ are the linear combinations that maximize the covariance (actually, the squared covariance, the sign is not important, just the direction).
  - S_{xx} and S_{yy} the covariance matrices, $\beta'\beta = 1$ and $a'S_{xx}a = 1$.

PLS

- We maximize the objective function:

$$\mathcal{L} = (\alpha' S_{xy} \beta)^2 - \gamma(\alpha'(S_{xx} + \lambda_x)\alpha - 1) - \mu(\beta'(S_{yy} + \lambda_y)\beta - 1).$$

- After some algebra we get the scores for X and Y, $t = Xw = Ew$ and $u = Yq = Fq$.
- Normalizing the scores $t = t/\sqrt{t't}$, and after simplifications and more algebra we get the loadings for X and Y: $p = E't$ and $q = F't$.

PLS

- Writing in matrix form w, t, p and q we get $R = W(P'W)^{-1}$ and finally
$$Y = XB + U, \text{ then } \hat{Y} = XB_{PLS},$$
where $B_{PLS} = R(T'T)^{-1}T'Y = RQ'$.

- For a nice introduction see P.H. Garthwaite (1994). An Interpretation of Partial Least Squares. JASA Vol 89, No 425, pp 122-127 and A. Hoskuldsson (1988). PLS Regression Methods. Journal of Chemometrics, Vol 2, pp 221-228.

  Note: Franses shows that if the B_{PLS} matrix has full rank it implies a different model for each of the columns of Y, and hence a model like $AR_{h,p}$. In the exceptional case that B_{PLS} has rank 1, then the $AR(p)$ appears.

Vector Autoregressive models and PLS

A $VAR(p)$ processes is defined as

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + CD_t + u_t,$$

where

- A_i, $i = 1, 2, \ldots, p$, the coefficient matrix.
- u_t a white noise process with variance covariance given by $\Sigma_u = E(u_t, u_t')$.
- C, matrix of regressor coefficients for deterministic factors
- D_t, vector of deterministic factors.

We also know that a $Var(p)$ can be written as a $Var(1)$ as follows

$$Y_t = AY_{t-1} + V_t,$$

with

$$Y_t = \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix}, \quad A = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{pmatrix}, \quad V_t = \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

If the eigenvalues of A are less than one, then the $VAR(p)$ is stable.

- We use the VAR representation to determine the order of the model.
- The general procedure is quite standard. Order $p = 0, \ldots, p_{\max}$ and choose the value of p that minimize some criteria. The criteria are usually written as:

$$IC(p) = \log|\hat{\Sigma(p)}| + C_T\varphi(K, p),$$

where
  - $\hat{\Sigma(p)} = T^{-1}\sum_{i=1}^{T}\hat{u}_t'\hat{u}_t$.
  - C_T an indexed sequence of the size T.
  - $\varphi(K, p)$ is a penalty function that involves the order of the $VAR(p)$.

- The most common information criteria are: Akaike (AIC), Schwarz-Bayesian (BIC), Hannan-Quinn (HQ) and Final Prediction Error (FPE):
  - Akaike: $AIC(p) = |\hat{\Sigma(p)}| + \frac{2}{t} pK^2$
  - Schwartz-Bayesian: $BIC(p) = |\hat{\Sigma(p)}| + \frac{\log T}{t} pK^2$
  - Hannan-Quinn: $HQ(p) = |\hat{\Sigma(p)}| + \frac{2 \log T}{t} pK^2$
  - Final Prediction Error: $FPE(p) = \left(\frac{T+p^*}{T-p^*}\right)^K \det(\hat{\Sigma(p)})$
- The AIC asymptotically overestimates the order of the model with a positive probability whereas BIC and HQ are consistent estimators of the order if the true value is less than or equal than p_{max}.

As for the univariate case, we can build the forecast with a recursive method:

$$y_{T+h|T} = A_1 y_{T+h-1|T} + \cdots + A_p y_{T+h-p|T} + C D_{T+h|T}$$

- We estimate A_i through LS

$$vec(\hat{A}) = \begin{pmatrix} \hat{A}_1 \\ \vdots \\ \hat{A}_p \end{pmatrix}.$$

- Under stationarity and ergodicity conditions for the VAR model (see Hamilton (1994), Lütkepohl (1991) among others), $vec(\hat{A})$ is consistent and asymptotically distributed with covariance matrix given by
$$\widehat{var}\left(vec(\hat{A})\right) = \hat{\Sigma} \otimes (Z'Z)^{-1},$$
where
$$\hat{\Sigma} = \frac{\sum_{t=1}^{T} \hat{\epsilon}_t' \hat{\epsilon}_t}{T - K}$$
and
$$\hat{\epsilon}_t = Y_t - \hat{A}' Z_t = Y_t - \hat{A}' Y_t$$
the LS residual at time t.

- The i-th element of $vec(\hat{A})$ is asymptotically normal (for a stable *VAR*) and the standard errors are the square roots of the diagonal elements of $\hat{\Sigma} \otimes (Z'Z)^{-1}$.
- The t-tests for the estimated coefficients are asymptotically correct.
- Other important situation for the *VAR* models is the presence of one or more unit roots for the y_j's. From the point of view of the Economic theory it means the study of a long run behavior plus the temporal dynamic of the series.

- **Cointegration:** The components of a k-dimensional vector y_t are cointegrated of order (d, b), denoted by $y_t \sim CI(d, b)$, if

    1. all components of y_t are $I(d)$.
    2. there exist a vector $\beta \neq 0$ such that $z_t = \beta' y_t \sim I(d-b)$, $b > 0$. The vector β is called the cointegration vector (Lütkepohl, 1991).

The $VAR(p)$ model can be written as a Transitory - Vector Error Correction Model ($VECM$)

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + CD_t + u_t,$$

where $\Gamma_i = -(A_{i+1} + \cdots + A_p)$, for $i = 1, \ldots, p-1$ and $\Pi = -(I - A_1 - A_2 - \cdots - A_p)$.

Or as a Long run - Vector Error Correction Model ($VECM$)

$$\Delta y_t = \Pi y_{t-p} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + CD_t + u_t,$$

where $\Gamma_i = -(I - A_1 - A_2 - \cdots - A_i)$, for $i = 1, \ldots, p-1$ and $\Pi = -(I - A_1 - A_2 - \cdots - A_p)$.

VAR

The matrix Π has the following characteristics:

1. If $rk(\Pi) = n$, the n linear combinations are stationary; in other words the *VECM* is no more than a *VAR* model (in levels).
2. If $rk(\Pi) = 0$, there is no linear combination that makes $\Pi y_{(t-1)}$ stationary, except for the trivial solution *i.e.*, it becomes a $VAR(p-1)$ in first differences.
3. If $0 < rk(\Pi) < n$, then $\Pi = \alpha\beta'$ (α and β with dimensions $n \times r$) and $\beta' y_{t-1}$ is stationary. Each column of β represent a long run relationship.

If the objective is to forecast series that are integrated or cointegrated, working with a VAR representation is quite appropriated (see Lütkepohl 2006)

VAR Example

- For the Mexican inflation example we specify the order of the model through the final error prediction criteria, it was $p = 2$.
- We also performed Johansen test to determine the presence of a long run relationship. Finding the following relationship (significant at a 1% level)

$$price + 9.68 - 0.43m - 0.89y + 0.1r = 0$$

- This relationship is congruent with the Economic theory behind it. The inflationary movement increases with monetary growth, the exceed on demand and with the reduction of the cost of money.

$$VAR - PLS(h, p)$$

- The VAR model will give us the DGP
- The VAR model will provide the autoregressive process that we will use to build the PLS regression
- We then form the matrices in a natural way as:

- For the X matrix we include the lag vector

$$X = Y_{t-1} = \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix},$$

- For Y, the observation until time t :

$$Y = Y_t = \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix},$$

using the X matrix with all the lags considered for the DGP.

- We can introduce exogenous variables with a C matrix, then

$$X = Y_{t-1}^{*} = \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \\ D_t \end{pmatrix},$$

and then, the matrix of coefficients

$$A^{*} = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p & C \\ I & 0 & \cdots & 0 & 0 & 0 \\ 0 & I & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \\ 0 & 0 & \cdots & I & 0 & 0 \end{pmatrix}$$

- Those are the basic ingredients for a $VAR - PLSX$ that will allow us to predict h steps ahead

$VAR - PLSX(h, p)$ Example

For the $VAR(p = 3) - PLSX(h = 24, j)$

- We kept 24 observation to have a long horizon of possible comparisons
- We consider dummies for the monthly effects
- For the optimal p we estimate the $VAR - PLSX$ model. We use R to fit

$$Y_{t,119\times 4} = X_{t,119\times 23} B_{23\times 4} + U_{t,119\times 4}$$

  and estimate and predict for the $VAR(3) - PLSX(24, j)$ in a recursive way as in the $VAR(p)$. (The last component agrees with the VAR(p) LS estimate)
- For the $pK + g = 23$ components and the 24 out of sample observations we get the MAPE to make the comparison

Prediction Interval: VAR-PLS

$VAR - PLS(h, p)$: Bootstrap

We also have to construct the prediction intervals for the VAR-PLS model.

We use a similar procedure to the one proposed by Pascual, Ruíz and Fresoli (2011). Bootstrap forecast of multivariate VAR models without using the backward representation. Working Paper 11-34, Statistics and Econometrics Series.
They use the seminal ideas of Kim (2001) and some results from a previous work, Pascual, L., J. Romo, and E. Ruiz (2004a). Bootstrap predictive inference for ARIMA processes, Journal of Time Series Analysis, 25, 449-465

For the VAR model they proposed a method that copes with:

1. The uncertainty given by the estimation of the parameter, building confidence regions using a bootstrap method
    - *This regions are valid under Gaussian assumptions (Lütkepolh et al, 1991), even though do not reflect, for small sample sizes, the asymmetric distribution of the predicted values (under estimated parameters)*
2. The backward representation makes calculations quite complicated, more in the case of the $VAR(p)$ representation and p taking values greater than 5 or 6, which are very common.
    - *"Pascual et al shows that the backward representation can be avoided without loosing the good properties of the bootstrap procedure"*

$VAR - PLS(h,p)$: Bootstrap

Of course we needed to adequate the procedure for the $VAR - PLS$ representation

1. Fit the model to get $Y_t = X_t \hat{B}_{PLS}$.
2. Obtain the standarized residuals $\hat{U}_t^*$ and the empirical distribution of the residuals
3. With the p initial values $Y_0 = \{Y_p, \ldots, Y_1\}$ and the results from Steps 1 and 2 generate Y_t^*, the bootstrap values, through the $\hat{U}_t^*$'s, which are independently drawn from its empirical distribution

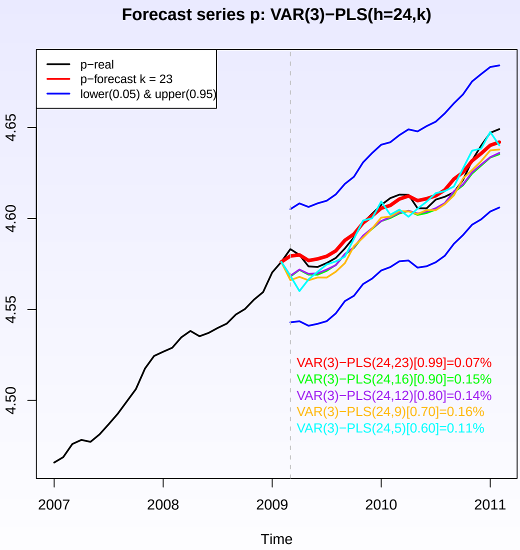
$$Y_t^* = X_t \hat{B}_{PLS} + \hat{U}_t^*, \quad t = 1, \ldots, n - p$$

4. Proceed in this way to get $\hat{Y}_{T+h}^*$, replicating steps 2 to 4 for $n = 1, \ldots, N$.
5. For each one of the n variables and the set of N forecast we get

$$CI_{T+h} = \{y_{n,T+k} | y_{n,T+k} \in [q_B^*(\tau), q_B^*(\tau - 1)]\},$$

where $q_B^*(\tau)$ is the τ-th quantile of $G_{n,B}^*(x) = \# \left(y_{n,T+k}^{*(b)} \leq x \right) / N$

VAR − PLS Example



VAR – *PLS* Example

- From the Economic point of view, the approximation is excellent.
- We observe that either using, $k = 23$ (0.99 of the variability) with an error percentage of 0.07% or with 70% explanation with an error percentage of 0.16%, the true value and the predicted one, for practical purposes, are almost identical.
- The bootstrap interval is very well behaved.

Note: The objective is to forecast Price, however since it is a multivariate model we also get forecasts for the other 3 variables with a forecast error (average of MAPE) of: 0.20% for the monetary base, 0.60% for the industry production index and 12.03% for the equilibrium interest rate.

$VAR(p)$ forecast

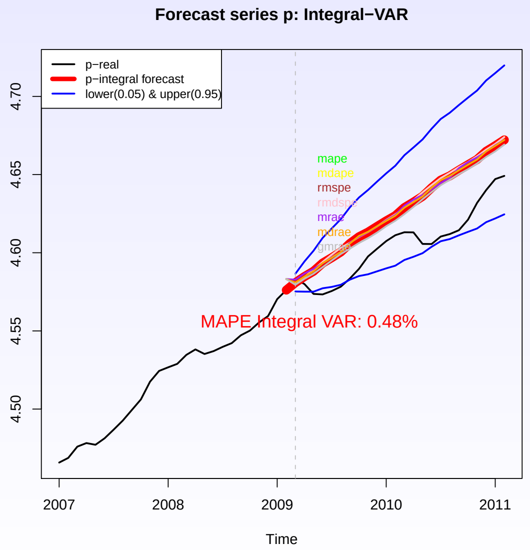
For the model $VAR(p=2)$

- Combining all the variables estimate our $VAR_j(p)$, $j = 1, 2, \ldots, 3696$
- For the 24 steps out of sample we use 7 different criteria to measure the behavior of the forecast (Hyndman & Koehler 2006):
  - MAPE: Mean Absolute Percentage Error
  - MdAPE: Median Absolute Percentage Error
  - RMSPE: Root Mean Square Percentage Error
  - RMdSPE: Root Median Square Percentage Error
  - MRAE: Mean Relative Absolute Error
  - MdRAE: Median Relative Absolute Error
  - GMRAE: Geometric Mean Absolute Error

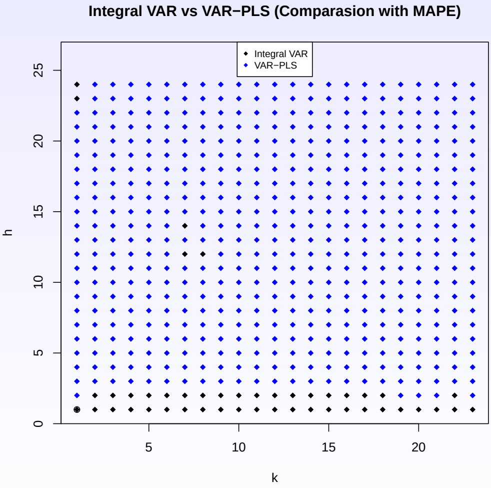
$VAR(p)$ forecast

- For the last 3 we need to work with a benchmark model (autoregressive order 1) and for $i = 1, \ldots, 24$ ($h = 24$) obtain the statistic:
$$test = \frac{Y_{t-i} - Y^f_{t+i,VAR_j(p)}}{Y_{t+i} - Y^f_{t+i,AR(1)}}$$
- With the 7 criteria, integrate the forecasts into one by taking their medians for the given horizon, repeat this idea for their upper and lower confidence limits to get the combined prediction interval

VAR – PLS Example



$VAR-PLS$ Example



Conclusions

In average, 91.67% of times, the PLS representation performed better than the integral VAR forecasts over all the components. It makes sense that the last components are less effective.

CONCLUSIONS

- The $VAR-PLS$ seems to be an attractive competitor against the integral VAR which is constructed for prediction purposes.
- One advantage is that the bootstrap prediction intervals include the uncertainty due to parameter estimation
- A second advantage is that the forecast reflects the trends and seasonal effects of the original series even for large number of steps ahead

Thanks !

References

- Bjorn-Helge Mevik and Ron Wehrens The pls Package: Principal Component and Partial Least Squares Regression in R, Journal of Statistical Software January 2007, Volume 18, Issue 2.
- Carreiro, Kapetorios and Marcellino (2010), Forecasting Government Bond Yields with large Bayesian VAR's. Working papers No. 662 School of Economics and Finance.
- Esposito Vinzi, V., W.W. Chin, J. Henseler and H. Wang (2007), Handbook of Partial Least Squares, Berlin: Springer.
- Philip Hans Franses (2006) Forecasting 1 to h steps ahead using partial least squares. Econometric Institute, Erasmus University Rotterdam, Econometric Institute Report 2006-47
- P.H. Garthwaite (1994). An Interpretation of Partial Least Squares. JASA Vol 89, No 425, pp122-127
- Hoskuldsson (1988). PLS Regression Methods. Journal of Chemometrics, Vol 2, pp 221-228

References

- Kim, J.H. (2001), Bootstrap after bootstrap prediction intervals for autoregressive models, Journal of Business & Economic Statistics, 19(1), 117-128.
- Lutkepolh, H. (1991), Introduction to Multiple Time Series Analysis, 2nd ed., Springer- Verlag, Berlin.
- Lutkepolh, H. (2006), Forecasting with VARMA models, in Elliot, G., C.W.J. Granger and Timmerman (eds.), Handbook of Economic Forecasting, Vol. 1, 287-325.
- McIntosh A.R., Bookstein F.L., Haxby J.V.and Grady C.L. (1996). Spatial Pattern Analysis of Functional Brain Images Using Partial Least Squares NeuroImage, Volume 3, Number 3, pp. 143-157(15). Academic Press
- Marcellino, Stock and Watson (2004), A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series.NBER
- Paul Nomikos and John F. MacGregor (1995) Multivariate SPC Charts for Monitoring Batch Processes Technometrics Vol. 37, No. 1, pp. 41-59

References

- Pascual, Ruiz y Fresoli (2011). Bootstrap forecast of multivariate VAR models without using the backward representation. Working Paper 11-34, Statistics and Econometrics Series,
- Pascual, L., J. Romo, and E. Ruiz (2004). Bootstrap predictive inference for ARIMA processes, Journal of Time Series Analysis, 25, 449-465
- Pesaran, H. H., Pick, A., Timmermann, A. (2010). Variable Selection and Inference for Multi-period Forecasting Problems. Journal of Econometrics
- Tiao, G.C., and D. Xu (1993), Robustness of maximum likelihood estimates for multi-step predictors: the exponential case, Biometrika, 80, 623-641

Acknowledgments: Partially supported by CONACYT CB No.105657; ECO2010-19357, Spain and FOMIX-NL C42 -178237

$VAR(p)$ Example

For the Integral-VAR, the optimum $VAR_J(P)$ are

Criterium	MAPE	MdAPE	RMSPE	RMdSPE
Statistic	0.16	0.12	0.19	0.12
Variable	r	r	r	r
Lags	3	2	3	2
Stationality	9	11	9	11
Specification	none	none	none	none

Criterium	MRAE	MdRAE	GMRAE
Statistic	0.16	0.12	0.11
Variable	r	r	r
Lags	2	3	5
Stationality	6	9	11
Specification	none	none	none