Var PLS

Graciela González Farías

Francisco Corona and Jesús Gonzalo
Centro de Investigación en Matemáticas, Campus Monterrey

ISBIS 2012

Bangkok, Thailand, June 17, 2012

Contenido

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

Motivación

- PLS is a technique that has bee proven its impact on many applications such as quality control starting with the Chemistry, batch processes, medical images analysis, microarrays, path modeling, classification, discrimination, spacio-temporal PLS models just to mention some, with authors such as McGregor, Nomikos, MacIntosh, V. Esposito Vinz, P. Garthwaite, and so on
- The method can be used in univariate and multivariate data as well
- *It has been shown that gives better prediction even when the standard assumptions are met*
- Phillip Hans Franses (2006) propose a methodology to construct the forecast h steps ahead in an optimal way, through an autoregresive order p: *An Autoregresive Partial Least Square denote as* $PLSAR(h, p)$

Our case of interest

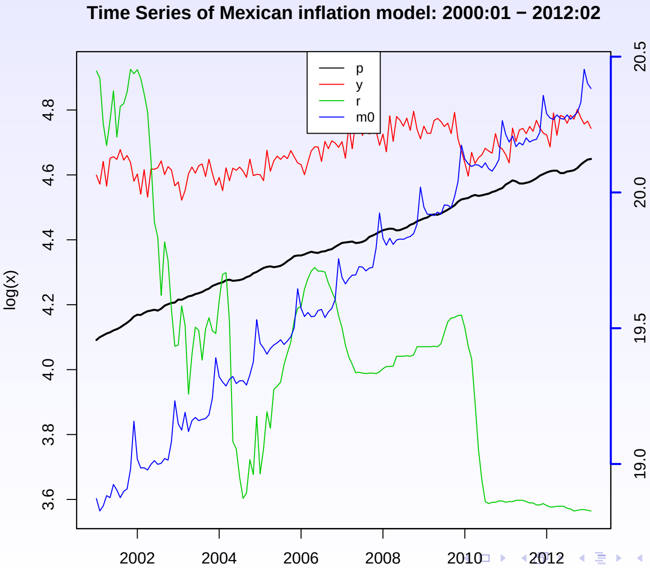
- Develop a model to predict the Mexican inflation, as precise as possible
- The model has to consider, as the principal source of the mexican inflation, the grow and the variation on the monetary condition of the country
- Irrespective of all possible discussions, there seems to be a common understanding to belived that the inflationary process, in the lung run, is a purely monetary fenomena
- Here we are not taking the discussion on the existence or not of such relationship but we will show its empirical properties with a model that is tested out of the sample via its error prediction measure

Our case of interest

We work with 4 indexes (we bulit those) from January 2000 to Feb 2012

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

Nuestro caso de interés



Our case of interest

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

    1. Give a multivariate representation based on the flexibility of the 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 a $VAR-PLS(h,p)$ and a forecast VAR model explicitly built for (integral predictor method)

$$PLSAR(h, p)$$

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

1- A single model for all horizons , an iterative procedure will come on hand

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

For the $AR(p)$ the classical procedure to get h step ahead forecast, plus the fact that we estimate the parameters (OLS).

$PLSAR(h, p)$

2- One model for each horizon, the variance will vary within each horizon and *one can count with 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

- OLS minimize the sum of square of ϵ_t 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 the forecast of an $AR(p)$ model quickly converge 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 exists adjacent correlation between the time series, and neither one of the above models take them 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 techinique very atractive to do so.

$PLSAR(h, p)$

3- Franses propose to arrange the information as

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

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

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

PLS

PLS

- PLS can be tract from different stand points, to us the relationship between its linear expression will be the best one, in order to related with a Vector Autoregresive 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 maximize the

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

under certain restrictions,

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

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

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

PLS

- We maximze the 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 y 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 y 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

$VAR - PLS$

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 deterministc factors
- D_t the vector of deterministic factors

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

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

con

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

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:

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

where

- $\Sigma(\hat{p}) = T^{-1}\sum_{i=1}^{T} \hat{u}_t'\hat{u}_t,$
- 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 criterias are: Akaike (AIC), Schwarz-Bayesiano (BIC), Hannan-Quinn (HQ) and Final predicction error (FPE):
  - Akaike: $AIC(p) = |\Sigma\hat{(p)}| + \frac{2}{t}pK^2$
  - Schwartz-Bayesiano: $BIC(p) = |\Sigma\hat{(p)}| + \frac{\log T}{t}pK^2$
  - Hannan-Quinn: $HQ(p) = |\Sigma\hat{(p)}| + \frac{2\log T}{t}pK^2$
  - Final prediction error: $FPE(p) = \left(\frac{T+p^*}{T-p^*}\right)^K \det(\Sigma\hat{(p)})$
- The AIC asymptotically overestimate the order of the model with a positive probability whereas BIC and HQ are consisten estimator of the order if, the real value of is less than or equal than p_max

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

$$y_{T+h\backslash T} = A_1 y_{T+h-1} + \cdots + A_p y_{T+h-p} + CD_{T+h}$$

- We estimate A_i throught OLS

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

- Under stationarty and ergodicity conditions for the VAR model (see Hamilton (1994), Lutkepohl (1991) among others), $vec(\hat{A})$ is consistent and asymtotically distributued 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 OLS residual at time t.

- El $i - th$ element of $vec(\hat{A})$ is asymptotically normal (for an a stable VAR) and the standard error are the square roots of the diagonal elements of $\hat{\Sigma} \otimes (Z'Z)^{-1}$.
- The $t - test$ 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'. From the point of view of the economic theory it mean the study of a lung 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 $y_t \sim$ 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 (Lutkepohl, 1991)

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

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

donde $\Gamma_i = -(I - A_1 - A_2 - \cdots - A_i)$, para $i = 1, \ldots, p-1$ y
$\Pi = -(I - A_1 - A_2 - \cdots - A_p)$.

$VAR-PLS$

The matrix Π has the following characteristics:

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

If the objective is to forecast series that are integrated or cointegrated working with a VAR representation is quite appropriated (see Lutkepohl 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 lung run relationship. Finding the following relationship which it was significant to 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 for the monetary grow, the exceed on demand and the reduction on the money cost

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

- The *VAR* model will give us the DGP
- The *VAR* model will provide the autoregressive process that we will use to built the *PLS* regression
- We built then the matrices in at 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 the 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 consider for the DGP.

$VAR - PLS(h, p)$

- 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 allows to predict h steps ahead

$VAR - PLS(h, p)$ Example

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

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

$VAR(p)$ forecast

Para el modelo $VAR(p = 3)$

- Combining all the variables estimate our $VAR_j(p)$, $j = 1, 2, \ldots, 3696$
- For the 24 steps out of the 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

- 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 forecast to one by taking the 0.5 quantile for the horizon consider, repeat it for their upper and lower confidence intervals to get the predicted integral interval

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 (2001and 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)

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

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

1. The uncertainty give by the estimation of the parameter, building confidence regions using their bootstrap method
   - *This regions are valid under Gaussian assumptions (Lütkepolh et al, 1991), even though do not reflect, for small sample size, the asymmetric distribution of the predicted values (under estimated parameters)*
2. The backward representation makes calculations quite complicate more in the case of the $VAR(p)$ representation and p taking values greater than 5 for example, which is 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 to the procedure to the $VAR - PLS$ representation

1. Fit the model to get $Y_t = X_t \hat{B}_{PLS}$.
2. Obtain the standarized residuals $\hat{U}^*_t$ y la and get the emprical distribution of the residuals
3. With the p initial values $Y_0 = \{Y_p, \ldots, Y_1\}$ and the values obtain in Step 1 and 2 generate Y^*_t, the bootstrap values, through were the $\hat{U}^*_t$ are independent drawn from its empirical distribution

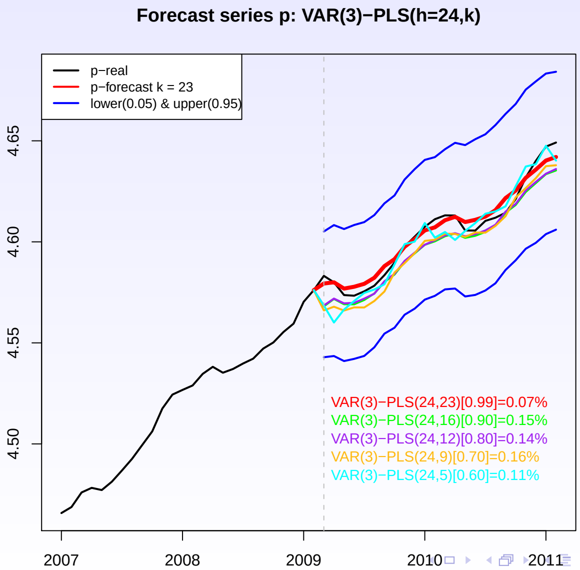
$$Y^*_t = X_t \hat{B}_{PLS} + \hat{U}^*_t, \quad t = 1, \ldots, n - p$$

4. We 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 $\tau-$th percentil of $G^*_{n,B}(x) = \# \left(y^{*(b)}_{n,T+k} \leq x \right) / N$

VAR – PLS Example



VAR – PLS Example

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

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

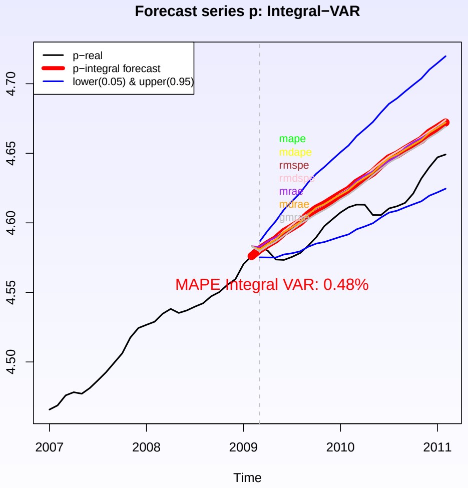
$VAR-PLS$ Example

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

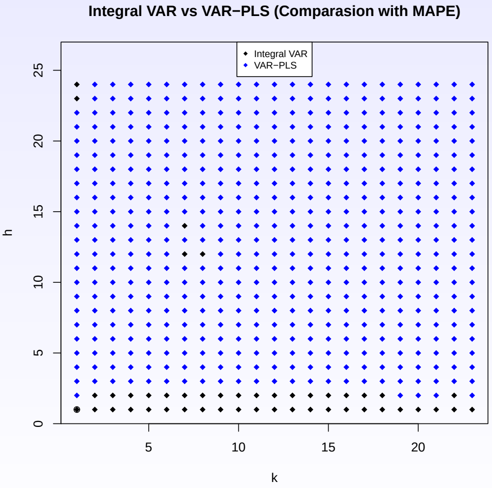
Criterio	MAPE	MdAPE	RMSPE	RMdSPE
Statistic	0.16	0.12	0.19	0.12
Variable	r	r	r	r
Lags	3	2	3	2
Stationality	9	11	9	11
Specification	none	none	none	none

Criterio	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

$VAR - PLS$ Example



VAR – PLS Example



In average 91.67% the PLS representation of the VAR over all the components. It make sense that the last components are less effective.

CONCLUSION

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

Thanks !

Reference

- 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. http://www.jstatsoft.org/
- Carreiro, Kapetorios and Marcellino (2010), Forecasting Government Bond Yields with large Bayesian VARs. 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

Reference

- 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

Reference

- 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