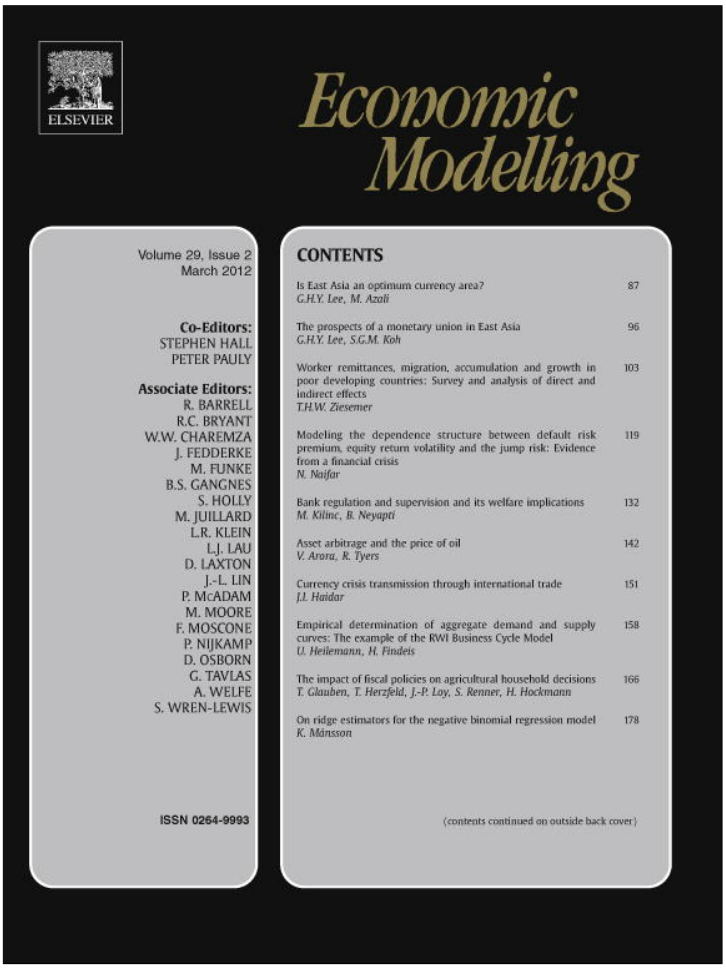

(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Economic Modelling 29 (2012) 1099–1105

Contents lists available at SciVerse ScienceDirect

Economic Modelling

journal homepage: www.elsevier.com/locate/ecmod

A medium-*N* approach to macroeconomic forecasting

Gianluca Cubadda [a,*], Barbara Guardabascio [b]

[a] Università di Roma "Tor Vergata", Rome, Italy
[b] ISTAT, Rome, Italy

ARTICLE INFO

Article history:
Accepted 24 March 2012
Available online xxxx

Keywords:
Partial least squares
Principal component regression
Dynamic factor models
Data-rich forecasting methods
Dimension-reduction techniques

ABSTRACT

This paper considers methods for forecasting macroeconomic time series in a framework where the number of predictors, *N*, is too large to apply traditional regression models but not sufficiently large to resort to statistical inference based on double asymptotics. Our interest is motivated by a body of empirical research suggesting that popular data-rich prediction methods perform best when *N* ranges from 20 to 40. In order to accomplish our goal, we resort to partial least squares and principal component regression to consistently estimate a stable dynamic regression model with many predictors as only the number of observations, *T*, diverges. We show both by simulations and empirical applications that the considered methods, especially partial least squares, compare well to models that are widely used in macroeconomic forecasting.

1. Introduction

Growing attention has recently been devoted to forecasting economic time series in a data rich framework (see, inter alia, Forni et al., 2005; Stock and Watson, 2002a). In principle, the availability of large data sets in macroeconomics provides the opportunity to use many more predictors than those that are conventionally used in typical small-scale time series models. However, exploiting this richer information set comes at the price of estimating a larger number of parameters, thus rendering numerically cumbersome or even impossible the application of traditional multiple regression models.

A standard solution to this problem is imposing a factor structure to the predictors, such that principal component [PC] techniques can be applied to extract a small number of components from a large set of variables. Some key results concerning forecasting with many predictors through the application of PCs are given in Stock and Watson (2002a, 2002b) and Forni et al. (2003, 2005). Recently, Gröen and Kapetanios (2008) have proposed partial least squares [PLS] as alternatives to PCs to extract the common factors. A different methodological framework is Bayesian regression as recently advocated by De Mol et al. (2008) and Banbura et al. (2010). Particularly, these authors attempted to solve the dimensionality problem by shrinking the forecasting model parameters using ridge regression [RR].

A common feature of the mentioned approaches is that statistical inference requires a double asymptotics framework, i.e. both the number of observations *T* and the number of predictors *N* need to diverge to ensure consistency of the estimators. However, an interesting question to be posed is how large the predictor set must be to improve forecasting performances. At the theoretical level, the answer provided by the double asymptotics method is clear-cut: the larger *N*, the smaller is the mean square forecasting error. However, Watson (2003) found that factor models offer no substantial predictive gain from increasing *N* beyond 50, Boivin and Ng (2006) showed that factors extracted from 40 carefully chosen series yield no less satisfactory results than using 147 series, Banbura et al. (2010) found that a vector autoregressive [VAR] model with 20 key macroeconomic indicators forecasts as well as a larger model of 131 variables, and Caggiano et al. (2011) documented that the best forecasts of the 7 largest European GDPs are obtained when factors are extracted from 12 to 22 variables only.

The above results advocate in favor of a sort of "medium-*N*" approach to macroeconomic forecasting. Specifically, we aim at solving prediction problems in macroeconomics where *N* is considerably larger than in typical small-scale forecasting models but not sufficiently large to resort to statistical inference that is based on double asymptotics methods. In order to accomplish this goal, we reconsider some previous results in the PLS literature in a time-series framework. Particularly, we argue that, under the so-called Helland and Almoy condition (Helland, 1990; Helland and Almoy, 1994), both principal component regression [PCR] and the PLS algorithm due to Wold (1985) provide estimates of a stable dynamic regression model that are consistent as *T* only diverges.

Since to date little is known on the statistical properties of PLS in finite samples, a Monte Carlo study is carried out to evaluate the forecasting performances of this method in a medium-*N* environment. To our knowledge, our simulation analysis is unique in that we simulate time series generated by stationary 20-dimensional VAR(2) processes that satisfy the Helland and Almoy condition. Indeed, several studies were devoted to compare PCR and PLS with other methods (see, inter alia, Almoy, 1996) but always in a static framework. Our results suggest that dynamic regression models estimated by PCR and, especially, PLS forecast well when compared to both OLS and RR.

* Corresponding author at: Dipartimento SEFEMEQ, Università di Roma "Tor Vergata", Via Columbia 2, 00133 Roma, Italy. Tel.: +39 06 72595847; fax: +39 06 2040219.
E-mail addresses: gianluca.cubadda@uniroma2.it (G. Cubadda), guardabascio@istat.it (B. Guardabascio).

0264-9993/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.econmod.2012.03.027

In the empirical application, we forecast four US macro time series by a rich variety of methods using similar variables as in the medium dimension VAR model by Banbura et al. (2010). The empirical findings indicate that PLS outperforms the competitors. Interestingly, Lin and Tsay (2006), Gröen and Kapetanios (2008) and Eickmeier and Ng (2011) reached similar conclusions using PLS as an alternative to PCs in large-*N* dynamic factor models.

The remainder of this paper is organized as follows. The main theoretical features of the suggested methods are detailed in Section 2. The Monte Carlo design and the simulation results are discussed in Section 3. Section 4 compares various forecasting procedures in empirical applications to US economic variables. Finally, Section 5 concludes.

2. Dynamic partial least squares and principal component regression

Let us suppose that the scalar time series to be forecasted, y_t, is generated by the following regression model

$$y_{t+1} = \beta' X_t + \varepsilon_{t+1}, \quad t = 1, \ldots, T, \tag{1}$$

where X_t is N-vector of stationary and ergodic time series, possibly including lags of y_{t+1}, ε_t is a serially uncorrelated error term with $E(\varepsilon_t) = 0$, $E\left(\varepsilon_t^2\right) = \sigma_\varepsilon^2$, $E\left(\varepsilon_t^4\right) < \infty$, and such that $E(\varepsilon_{t+1}|X_t) = 0$. Moreover, we assume that deterministic elements are absent from both time series y_t and X_t, and that each element of X_t has unit variance.

In order to reduce the number of parameters to be estimated in model (1), we follow Helland (1990) and Helland and Almoy (1994) and take the following condition:

Condition 1. *(Helland and Almoy) Let* $\mathrm{E}(X_t y_{t+1}) = \Sigma_{xy}$ *and* $\mathrm{E}(X_t X'_t) = \Sigma_{xx} = \Upsilon \Lambda \Upsilon'$, *where* Υ *is the eigenvector matrix of* Σ_{xx} *and* Λ *is the associated diagonal eigenvalue matrix. We assume that*

$$\Sigma_{xy} = \Upsilon_q \xi, \tag{2}$$

where Υ_q *is a matrix formed by q eigenvectors (not necessarily those associated with the q largest eigenvalues) of* Σ_{xx}, *and* ξ *is a q-vector with all the elements different from zero.*

The above condition is discussed at length in Helland (1990) and Næs and Helland (1993). Essentially, it is equivalent to require that the predictors X_t can be decomposed as

$$X_t = \theta R_t + \theta_\perp E_t,$$

where $R_t = \theta' X_t$, $E_t = \theta'_\perp X_t$, θ and $\theta_\perp$ are, respectively, orthonormal matrices of dimension $N \times q$ and $N \times (N-q)$ such that $\theta\theta' = I_N - \theta_\perp \theta'_\perp$, $E(R_t E'_t) = 0$, and $\Sigma_{xy} = \theta E(R_t y_{t+1})$. R_t and E_t are, respectively, called the relevant and irrelevant components of predictors X_t. The linear combinations $\Upsilon'_q X_t$ that span the space of the relevant components are then called the relevant principal components.

In principle, Condition 1 is in line with the common view that macroeconomic time series are mainly led by few aggregate shocks (e.g. demand and supply shock), which are independent from minor causes of variability (e.g. errors in variables or sector-specific shocks). In Section 4 we will tackle this issue from an empirical viewpoint.

Notice that Condition 1 implies

$$\beta = \Upsilon_q \Lambda_q^{-1} \xi, \tag{3}$$

where Λ_q is the diagonal eigenvalue matrix associated with Υ_q. Hence, model (1) has the following factor structure:

$$y_{t+1} = \xi' F_t + \varepsilon_{t+1},$$

where $F_t = \Lambda_q^{-1} \Upsilon'_q X_t$. Hence, since $\mathrm{E}(y_{t+1}|X_t)$ is a linear transformation of F_t, the predictable component of y_{t+1} is entirely captured by the q components F_t. This is not necessarily the case in dynamic factor models, where the idiosyncratic term is generally not an innovation.[1]

At the population level, PCR computes the prediction for y_{t+1} as $\beta'_{PCR} X_t$ where

$$\beta_{PCR} = \Upsilon_q \Lambda_q^{-1} \Upsilon'_q \Sigma_{xy}. \tag{4}$$

In view of Eq. (3), it is clear under Condition 1 that we have $\beta_{PCR} = \beta$. However, in empirical applications the relevant principal components must be selected and the eigenvalues of the sample covariance matrix of the predictors offer no guidance on this choice. Indeed, Condition 1 does not impose that the eigenvalues associated to the eigenvectors Υ_q are the q largest ones of matrix Σ_{xx} and there is no sound theoretical reason why this should occur (see, inter alia, Hadi and Ling, 1998). As shown by Helland (1990), PLS offer an effective way to overcome this problem.

PLS, introduced by Wold (1985), is an iterative procedure that aims at maximizing the covariance between a target variable and linear combinations of its predictors. In order to accomplish this goal, the first PLS component $\omega'_1 X_t$ is built such that the weights ω_1 are equal to the covariances between the predictors X_t and the target variable y_{t+1}. The second PLS component $\omega'_2 X_t$ is similarly constructed using a new target variable that is obtained by removing the linear effect of the first component on y_{t+1}. In general, the weights of the subsequent PLS factors are set equal to the covariances between X_t and a novel target variable that is obtained by removing the linear effects of all the previously obtained PLS components on y_{t+1}. Hence, let $\beta'_{PLS} X_t$ indicate the prediction of y_{t+1} using the first q PLS components, where

$$\beta_{PLS} \equiv \Omega_q \left(\Omega'_q \Sigma_{xx} \Omega_q\right)^{-1} \Omega'_q \Sigma_{xy}, \tag{5}$$

$\Omega_q = (\omega_1, \ldots, \omega_q)$, and

$$\omega_{i+1} = \Sigma_{xy} - \Sigma_{xx} \Omega_i (\Omega'_i \Sigma_{xx} \Omega_i)^{-1} \Omega'_i \Sigma_{xy}, \quad i = 1, \ldots, N-1 \tag{6}$$

with $\omega_1 = \Sigma_{xy}$. Since it follows by induction from Eq. (6) that Ω_q lies in the space spanned by the eigenvectors Υ_q, it is easy to see that $\omega_i = 0$ for $i = q+1, \ldots, N$ and $\beta_{PLS} = \beta$.

Further features of PLS are better understood by considering the following equivalent way to obtain the weights Ω_q (Helland, 1990). Let us define $V_{0,t} = X_t$ and

$$V_{i,t} = V_{i-1,t} - \phi_i f_{i,t} = X_t - \sum_{j=1}^{i} \phi_j f_{j,t}, \quad i = 1, \ldots, q, \tag{7}$$

where $\omega_i = \mathrm{E}\left(V_{i-1,t} y_{t+1}\right)$, $\phi_i = \mathrm{E}\left(f_{i,t} V_{i-1,t}\right) / \mathrm{E}\left(f'_{i,t} f_{i,t}\right)$, and $f_{i,t} = \omega'_i V_{i-1,t}$ is the i-th PLS factor.

Eq. (7) tells us that the i-th PLS factor $f_{i,t}$ is constructed as a linear combination of the predictors X_t (with weights equal to elements of ω_i) after having removed the linear effects of the previously constructed factors $f_{1,t}, \ldots f_{i-1,t}$. Moreover, by premultiplying each side of Eq. (7) by ω'_i, we see that i-th PLS component can be rewritten as

$$\omega'_i X_t = f_{i,t} + \omega'_i \sum_{j=1}^{i-1} \phi_j f_{j,t}.$$

[1] This property is shared with models obtained through the reduced-rank VAR methodology, see, inter alia, Centoni et al. (2007). However, reduced-rank regression requires the specification of the multivariate model for series (y_t, X'_t) and it is not appropriate for a medium N framework, see inter alia Cubadda and Hecq (2011).

Putting these two observations together, we conclude that the PLS factors $(f_{1,t},\dots f_{q,t})'$ are uncorrelated with one other and that they are a non-singular linear transformation of the PLS components $\Omega_q'X_t$. Hence, $\beta_{PLS}'X_t$ may be equivalently obtained by a linear regression of y_{t+1} on the PLS factors.

The above alternative way of deriving PLS, which essentially is the population version of the algorithm popularized by Wold (1985), reveals that the PLS factors are orthogonal linear combinations of predictors X_t that are obtained by maximizing their covariances with the target variable y_{t+1}. Hence, differently from the PCs, the PLS factors take into account of the comovements between the target series and the predictors.

Since both PCR and PLS are continuous functions of the elements of the variance–covariance matrix of $(y_{t+1},X'_t)'$, it follows that under Condition 1 the sample versions of Eqs. (4) and (5) are consistent estimators of β as $T\to\infty$ by the consistency of the sample variance–covariance matrix of stationary and ergodic processes and the continuous mapping theorem. Chun and Keleş (2010) have recently proved the consistency of Eq. (5) when $N/T\to 0$ as T diverges. This allows for the number of predictors to increase but at a slower rate than the sample size, which makes PLS well suitable for a medium N framework.

However, apart from consistency, little is known so far regarding the statistical properties of PCR and PLS. This seriously limits the use of formal testing procedures to choose q in empirical applications. Hence, this choice is usually carried out by cross-validation selection rules. We will deal with this issue in Section 4.

In comparative terms, Helland and Almoy (1994) analyzed the expected prediction errors of PCR and PLS and concluded that no method asymptotically dominates the other. Stoica and Söderström (1998) proved that, under mild conditions, these methods are equivalent to within a first-order approximation. In the next sections we will assess the forecasting performances of PCR, PLS and other methods both by simulations and empirical examples.

3. Monte Carlo analysis

Several simulation analyses have compared the performances of PLS, PCR and other methods in a static framework (see, inter alia, Almoy, 1996). However, there is a lack of an extensive Monte Carlo study that examines the forecasting performances of PCR and PLS under the Helland and Almoy condition. In order to fill this gap, we resort to a sort of reverse engineering approach. First, we generate the relevant components from r stationary AR(2) processes and the irrelevant ones from $(n–r)$ white noise processes, which are independent on each other. Second, we construct an n-dimensional VAR(2) process by taking orthogonal linear transformations of the previously obtained components. By construction, each element of this VAR(2) process follows a stable dynamic regression model that satisfies the Helland and Almoy condition.

We start by simulating the following n-vector of stationary time series

$$H_t = \alpha + \Pi_1 H_{t-1} + \Pi_2 H_{t-2} + \epsilon_t,$$

where Π_2 is a diagonal matrix with the first r diagonal elements π_2 drawn from a $U_n[-0.95,0.95]$ and the remaining elements equal to zero, Π_1 is a diagonal matrix with the first r diagonal elements π_1 are from a $U_n[\pi_2-1,1-\pi_2]$ and the remaining elements equal to zero, α is n-vector of constant terms that are drawn from a $U_n[0,n]$, and ϵ_t are i.i.d. $N_n(0,I_n)$.

Moreover, we take the following linear transformation of the series H_t

$$Y_t = QH_t = Q\alpha + Q\Pi_1 Q'Y_{t-1} + Q\Pi_2 Q'Y_{t-2} + Q\epsilon_t, \quad (8)$$

where Q is an orthogonal matrix that is obtained by the QR factorization of a $n\times n$-matrix such that its columns are generated by n i.i.d. $N_n(0,I_n)$.

We notice from Eq. (8) that each element of Y_t is generated by a stable dynamic regression model with the same form as Eq. (1), where y_t is a generic element of the vector series Y_t, ε_t is the corresponding element of Q_t, and $X_t=[Y_t',Y_{t-1}']'$. Since the relevant and irrelevant components of X_t are respectively given by

$$R_t = [Y'_{t-1}Q_{\cdot r}, Y'_{t-2}Q_{\cdot r}]'$$

and

$$E_t = [Y'_{t-1}Q_{\cdot n-r}, Y'_{t-2}Q_{\cdot n-r}]',$$

where $[Q_{\cdot r},Q_{\cdot n-r}]=Q$, and $Q_{\cdot r}$ is an $n\times r$-matrix, we conclude that Condition 1 is satisfied.

Remarkably, Eq. (8) unravels that series Y_t are generated from VAR processes with a reduced-rank structure. Hence, there is a close link between the Helland and Almoy condition and reduced-rank VAR models, which are commonly used in macroeconometrics because of their statistical and economic properties; see, inter alia, Cubadda (2007) and Cubadda et al. (2009).

We compare four direct forecasting methods. The first one is the h-step ahead OLS forecast of $y_{\tau+h}$, for $\tau=T,\dots,T+T^*-h$ which is obtained as $X_\tau'\hat\beta^h$ where $\hat\beta^h=(X'X)^{-1}X'y$, $X=[X_1,\dots,X_{T-h}]'$, and $y=[y_{h+1},\dots,y_T]'$.

The second method is the ridge regression [RR] forecast, as suggested by De Mol et al. (2008). Particularly, the RR forecast of $y_{\tau+h}$ is obtained as $X_\tau'\hat\beta^h_\lambda$ where

$$\hat\beta^h_\lambda = (X'X+\lambda I_n)^{-1}X'y,$$

and λ is a shrinkage scalar parameter. Since De Mol et al. (2008) documented that superior forecasting performances are obtained for values of λ between half and ten times the number of predictors N, we use $\lambda/N=0.5,1,2,5,10$.

The third method is the h-step ahead PCR forecast of $y_{\tau+h}$, which is obtained as $X_\tau'\hat\beta^h_{PCR}$ where

$$\hat\beta^h_{PCR} = \hat\Upsilon_q\hat\Lambda_q^{-1}\hat\Upsilon_q'X'y,$$

and $X\hat\Upsilon_q$ are the q sample PCs that are most correlated with y.

Finally, the last method is the h-step ahead PLS forecast of $y_{\tau+h}$, which is obtained as

$$X'_\tau\hat\beta^h_{PLS} = \hat F'_\tau\left(\hat F'\hat F\right)^{-1}\hat F'y,$$

where $\hat F=\left(\hat F_1,\dots,\hat F_{T-h}\right)'$, and $\hat F_t=\left(\hat f_{1,t},\dots\hat f_{q,t}\right)'$ are obtained recursively from Eq. (7) having substituted the population covariances with their sample analogs.

We evaluate the competing methods by means of the mean square forecast error [MSFE] relative to an AR(2) forecast. To construct these relative MSFEs, we simulate systems of $n=10,20$ variables, i.e. $N=2n$ predictors, with $r=1,2,3$, i.e. $q^*=2r$ relevant components. Having generated $T+170$ observations of the vector series Y_t, the first 50 points are used as a burn-in period, the last $T^*=120$ observations are used to compute the h-step ahead forecast errors and the intermediate T observations are used to estimate the various models.

When $n=10$ (20), we set $T=80,120$ (240,360) corresponding to 20,30 years of quarterly (monthly) observations, and the relative h-step ahead MSFEs are computed for $h=1,2,4,8$ (1,3,6,12). Since the literature is relatively silent on the choice of q, we examine the performances of PLS and PCR for $q=1,2,4,6,8$ despite the true number of relevant component is $q^*=2,4,6$. This will allow us to investigate the

Table 1
Simulations, relative MSFE.

$N=20, T=80$												
	$q^*=2$				$q^*=4$				$q^*=6$			
Models	$h=1$	$h=2$	$h=4$	$h=8$	$h=1$	$h=2$	$h=4$	$h=8$	$h=1$	$h=2$	$h=4$	$h=8$
PLS(1)	1.035	0.984	**0.975**	**0.980**	0.966	1.007	0.977	**0.976**	0.959	1.074	1.018	0.995
PLS(2)	1.129	**0.971**	0.978	0.988	1.016	0.966	**0.967**	0.980	0.944	1.001	0.988	**0.988**
PLS(4)	1.203	0.988	1.004	1.021	1.091	0.955	0.980	1.008	1.008	0.950	**0.976**	1.004
PLS(6)	1.213	1.017	1.038	1.059	1.105	0.968	1.005	1.043	1.026	0.952	0.995	1.038
PLS(8)	1.214	1.050	1.070	1.098	1.106	0.991	1.036	1.080	1.029	0.968	1.021	1.073
PCR(1)	1.012	0.986	0.980	0.982	1.002	1.004	0.986	0.979	1.054	1.068	1.028	0.998
PCR(2)	1.050	0.989	0.988	0.997	1.002	0.985	0.978	0.987	0.991	1.000	0.992	0.993
PCR(4)	1.107	1.010	1.015	1.031	1.036	0.990	0.997	1.019	0.990	0.981	0.992	1.016
PCR(6)	1.146	1.038	1.046	1.068	1.059	1.005	1.022	1.055	1.000	0.985	1.011	1.048
PCR(8)	1.172	1.064	1.077	1.104	1.077	1.023	1.050	1.089	1.011	0.997	1.038	1.084
OLS	1.214	1.245	1.283	1.341	1.106	1.157	1.239	1.319	1.029	1.105	1.215	1.318
RR(0.5)	1.105	1.129	1.155	1.190	1.004	1.047	1.108	1.161	0.933	0.998	1.082	1.148
RR(1)	1.055	1.076	1.097	1.123	0.964	1.003	1.054	1.098	0.903	0.962	1.033	1.087
RR(2)	1.007	1.024	1.042	1.060	0.933	0.966	1.007	1.041	**0.887**	**0.938**	0.993	1.034
RR(5)	**0.972**	0.983	0.994	1.005	**0.929**	**0.952**	0.976	0.996	0.914	0.950	0.977	0.996
RR(10)	0.978	0.983	0.987	0.991	0.971	0.982	0.988	0.993	0.986	1.005	1.002	0.999

MSFE are relative to an AR(2) forecast. $RR(\lambda/N)$ indicates RR with a shrinking parameter λ. Bold figures indicate the best forecasting method.

Table 2
Simulations, relative MSFE — VAR(10).

$N=20, T=120$												
	$q^*=2$				$q^*=4$				$q^*=6$			
Models	$h=1$	$h=2$	$h=4$	$h=8$	$h=1$	$h=2$	$h=4$	$h=8$	$h=1$	$h=2$	$h=4$	$h=8$
PLS(1)	1.002	0.983	0.974	**0.981**	0.947	1.016	0.986	0.984	0.919	1.072	1.013	0.986
PLS(2)	1.058	**0.964**	**0.970**	**0.981**	0.966	0.965	0.966	0.979	0.890	0.992	0.978	0.973
PLS(4)	1.087	0.966	0.982	0.997	0.998	**0.940**	**0.964**	0.989	0.909	0.922	**0.951**	**0.972**
PLS(6)	1.088	0.982	1.000	1.019	0.999	0.945	0.977	1.008	0.910	0.912	0.953	0.987
PLS(8)	1.088	0.999	1.020	1.039	0.999	0.957	0.992	1.028	0.910	0.916	0.964	1.005
PCR(1)	0.986	0.979	0.975	**0.981**	0.985	1.001	0.986	0.980	1.031	1.063	1.017	0.985
PCR(2)	1.000	0.974	0.974	0.986	0.961	0.968	0.967	**0.978**	0.933	0.976	0.968	0.968
PCR(4)	1.031	0.983	0.988	1.004	0.971	0.963	0.973	0.995	0.911	0.945	0.959	0.977
PCR(6)	1.051	0.997	1.006	1.025	0.982	0.969	0.986	1.014	0.909	0.941	0.967	0.994
PCR(8)	1.065	1.011	1.024	1.045	0.988	0.976	1.000	1.034	0.909	0.940	0.977	1.010
OLS	1.088	1.108	1.140	1.173	0.999	1.045	1.103	1.156	0.910	0.978	1.062	1.128
RR(0.5)	1.047	1.066	1.094	1.121	0.959	1.002	1.054	1.099	0.875	0.939	1.014	1.068
RR(1)	1.023	1.040	1.066	1.089	0.940	0.980	1.027	1.068	0.861	0.922	0.991	1.039
RR(2)	0.994	1.010	1.033	1.052	0.921	0.958	0.999	1.033	**0.852**	**0.909**	0.968	1.008
RR(5)	**0.970**	0.982	0.998	1.011	**0.917**	0.947	0.975	1.000	0.871	0.918	0.958	0.983
RR(10)	0.982	0.987	0.995	1.000	0.947	0.970	0.983	0.995	0.933	0.969	0.984	0.991

See the notes for Table 1.

implications on forecasting of misspecifying the number of factors for both PCR and PLS. The results, reported in Tables 1–4, are based on 5000 replications of series y_t.

The results indicate that OLS is generally outperformed by the competitors. Indeed, OLS performs similarly as the other methods only when $q^*=6$ and with the largest T for each n. This finding suggests that the cost of ignoring restrictions on β given by Eq. (2) is high in a medium-N framework, even when the sample size is large.

Looking at the alternative methods, PLS is overall the best performer. However, RR always produces the most accurate 1-step forecasts and it performs similarly as PLS when T is small, whereas PCR performs best for the largest forecast horizon when T is large.

Regarding the implications of the choice of the number of factors, the best results for both PLS and PCR are generally obtained with q equal or slightly smaller than q^* except for $h=1$, for which a model with a lesser number of components than q^* produces a lowest MSFE in most cases. However, for both PLS and PCR, the forecasting performances appear to be reasonably robust to the choice of q. In contrast, the performance of RR depends crucially on the choice of the shrinking parameter λ. In general, λ should increase as q^* gets smaller and N gets larger. Finally, the methods that appear to benefit more from a larger sample size are OLS and PLS.

Overall, PLS appears to be a valid alternative to more well-known forecasting methods in a medium-N framework, at least when Condition 1 is satisfied. In the next section, we evaluate the relative merits of PCR and PLS in an empirical exercise.

4. Empirical application

In order to perform our empirical out-of-sample forecasting exercise, we use a similar data-set as Banbura et al. (2010) for their medium dimension VAR model. It consists of 19 US monthly time series divided in three groups: i) real variables such as Industrial Production, employment; ii) asset prices such as stock prices and exchange rates; iii) nominal variables such as consumer and producer price indices, wages, money aggregates. The variables are listed in Table 5, along with the transformations that we apply in order to render them stationary, thus obtaining the vector series Y_t.[2] The data are observed at the monthly frequency for the period

[2] Due to the impossibility of testing for cointegration in a 19-dimensional VAR model, we remove unit roots from the data by properly differencing each individual series.

Table 3
Simulations, relative MSFE — VAR(20).

$N=40$, $T=240$												
	$q^*=2$				$q^*=4$				$q^*=6$			
Models	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
PLS(1)	1.019	0.970	0.980	**0.984**	0.961	0.958	0.975	**0.978**	0.926	0.955	0.976	0.979
PLS(2)	1.080	**0.968**	**0.978**	0.987	1.000	**0.944**	**0.966**	**0.978**	0.941	0.933	0.963	**0.976**
PLS(4)	1.118	0.975	0.985	0.996	1.046	0.945	0.967	0.984	0.986	**0.925**	**0.957**	0.979
PLS(6)	1.120	0.984	0.994	1.006	1.050	0.951	0.974	0.993	0.992	0.929	0.961	0.987
PLS(8)	1.120	0.994	1.004	1.017	1.051	0.959	0.983	1.003	0.992	0.934	0.968	0.996
PCR(1)	0.978	0.974	**0.978**	0.985	0.959	0.968	0.974	0.980	0.953	0.969	0.973	0.980
PCR(2)	0.991	0.973	0.980	0.988	0.949	0.954	0.967	**0.978**	0.926	0.946	0.961	0.977
PCR(4)	1.019	0.981	0.989	0.998	0.967	0.956	0.972	0.986	0.931	0.943	0.963	0.982
PCR(6)	1.040	0.990	0.999	1.008	0.984	0.962	0.980	0.996	0.942	0.945	0.969	0.991
PCR(8)	1.056	0.998	1.008	1.019	0.998	0.968	0.988	1.005	0.951	0.950	0.977	1.000
OLS	1.120	1.149	1.167	1.190	1.051	1.100	1.136	1.169	0.992	1.065	1.118	1.161
RR(0.5)	1.079	1.106	1.122	1.140	1.010	1.057	1.090	1.118	0.953	1.022	1.070	1.107
RR(1)	1.052	1.078	1.093	1.109	0.986	1.031	1.062	1.088	0.931	0.998	1.042	1.076
RR(2)	1.020	1.043	1.057	1.070	0.959	1.001	1.029	1.052	0.909	0.970	1.011	1.040
RR(5)	0.982	1.001	1.012	1.023	0.932	0.968	0.991	1.010	**0.892**	0.943	0.976	1.001
RR(10)	**0.967**	0.982	0.992	1.001	**0.930**	0.959	0.978	0.994	0.899	0.939	0.965	0.985

See the notes for Table 1.

Table 4
Simulations, relative MSFE.

$N=40$, $T=360$												
	$q^*=2$				$q^*=4$				$q^*=6$			
Models	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
PLS(1)	0.995	0.973	0.981	0.985	0.953	0.968	0.988	0.993	0.916	0.956	0.976	0.978
PLS(2)	1.031	**0.965**	**0.977**	**0.984**	0.970	0.949	0.971	**0.986**	0.912	0.928	0.956	0.973
PLS(4)	1.047	0.967	0.980	0.989	0.991	**0.944**	**0.967**	0.989	0.936	**0.917**	**0.947**	0.971
PLS(6)	1.047	0.972	0.986	0.995	0.991	0.946	0.971	0.995	0.938	**0.917**	0.950	0.976
PLS(8)	1.047	0.978	0.992	1.002	0.991	0.951	0.976	1.001	0.937	0.920	0.954	0.982
PCR(1)	0.969	0.975	0.979	0.985	0.958	0.974	0.980	0.988	0.952	0.969	0.973	0.976
PCR(2)	0.972	0.971	0.978	0.985	0.941	0.958	0.971	**0.986**	0.912	0.940	0.954	**0.970**
PCR(4)	0.987	0.973	0.983	0.991	0.947	0.955	0.972	0.991	0.907	0.932	0.951	0.973
PCR(6)	0.999	0.978	0.989	0.997	0.956	0.958	0.975	0.997	0.912	0.932	0.955	0.979
PCR(8)	1.009	0.983	0.994	1.003	0.963	0.961	0.980	1.003	0.917	0.934	0.958	0.985
OLS	1.047	1.073	1.089	1.104	0.991	1.038	1.069	1.105	0.937	0.999	1.042	1.084
RR(0.5)	1.031	1.056	1.071	1.084	0.975	1.021	1.050	1.083	0.920	0.981	1.021	1.061
RR(1)	1.018	1.043	1.057	1.070	0.964	1.009	1.037	1.068	0.909	0.969	1.008	1.046
RR(2)	1.001	1.025	1.038	1.050	0.949	0.993	1.019	1.048	0.896	0.954	0.990	1.025
RR(5)	0.976	0.997	1.009	1.020	0.931	0.970	0.993	1.018	**0.883**	0.935	0.966	0.997
RR(10)	**0.964**	0.982	0.992	1.002	**0.926**	0.960	0.980	1.000	0.885	0.930	0.956	0.982

See the notes for Table 1.

Table 5
Data description.

Code	Variable	Transformation
W875RX1	Personal income	$(1-L)ln$
PCECC96	Real consumption	$(1-L)ln$
INDPRO	Industrial production index	$(1-L)ln$
TCU	Capacity utilization	$(1-L)$
UNRATE	Unemployment rate	$(1-L)$
PAYNSA	Employees on nonfarm payrolls private	$(1-L)ln$
HOUST	Housing starts	ln
M1SL	Money stock: M1	$(1-L^{12})(1-L)ln$
M2SL	Money stock: M2	$(1-L^{12})(1-L)ln$
TOTADJRES	Institutional total adjusted reserves	$(1-L^{12})(1-L)ln$
BOGNONBR	Nonborrowed reserves	$(1-L^{12})(1-L)ln$
SP500	Standard and poor's common stock price index	$(1-L)ln$
FEDFUNDS	Effective federal funds interest rate	$(1-L)$
GS1	U.S. treasury interest rate	$(1-L)$
NEER	U.S. effective exchange rate	$(1-L)ln$
PPIFGS	Producer price index: finished goods	$(1-L^{12})(1-L)ln$
CPIAUCNS	Consumer price index	$(1-L^{12})(1-L)ln$
PCECTPI	Personal consumption expenditure deflator	$(1-L^{12})(1-L)ln$
AHETP	Average hourly earnings	$(1-L^{12})(1-L)ln$

1959.01–2007.12. Finally, the variables to be forecasted are Industrial Production (IP), Employment (EMP), Federal Funds Rate (FYFF), and Consumer Price Index (CPI).

For all the competing methods, the four target series are

$$y^h_{t+h} = \left(1-L^h\right) ln(\mathtt{IPI}_{t+h}),$$
$$y^h_{t+h} = \left(1-L^h\right) ln(\mathtt{EMP}_{t+h}),$$
$$y^h_{t+h} = \left(1-L^h\right) \mathtt{FYFF}_{t+h},$$
$$y^h_{t+h} = \left(1-L^h\right)\left(1-L^{12}\right) ln(\mathtt{CPI}_{t+h}),$$

for $h=1,3,6,12$, and the predictors are $X_t=[Y_t,..,Y_{t-p+1}]'$. Along with PLS, PCR and RR, we consider two additional approaches coming from the large-N literature. The first one, labeled as SW, is the Stock and Watson (2002a, 2002b) dynamic factor model, which computes the h-step ahead forecast of $y^h_{\tau+h}$ as $W'_\tau \hat{\beta}^h_{SW}$, where $W_t = \left[Z'_\tau \hat{\Psi}_q, Y^{L'}_t\right]'$, $Y_t=[y_t,Z'_t]'$, $Y^L_t=[y_t,..,y_{t-p+1}]'$, $\hat{\Psi}_q$ are the eigenvectors associated with the q largest eigenvalues of $Z'Z$, and $Z=[Z_1,\ldots,Z_T]'$.

Table 6
IPI, relative MSFE.

Models	$h=1$	$h=3$	$h=6$	$h=12$
Sample: 1975–2007				
PLS	**0.849**	**0.824***	**0.912*****	**0.931****
GK	0.952	0.959	1.120	1.199
PCR	0.949	1.048	1.221	1.419
SW	0.904	1.083	1.217	1.406
RR	0.983	1.034	1.185	1.372
Sample: 1975–1984				
PLS	**0.744**	**0.731**	**0.827****	**0.841**
GK	0.853	0.871	1.042	1.075
PCR	0.874	0.973	1.130	1.275
SW	0.890	1.072	1.155	1.263
RR	0.880	0.937	1.058	1.161
Sample: 1985–2007				
PLS	0.971	**0.968**	**1.031***	**1.046*****
GK	1.068	1.096	1.231	1.357
PCR	1.035	1.165	1.351	1.603
SW	**0.921**	1.100	1.305	1.587
RR	1.103	1.185	1.364	1.641

PLS forecasts are obtained using $p=11$, $q=2$; GK forecasts are obtained using $p=8$, $q=4$; PCR forecasts are obtained using $p=2$, $q=7$; SW forecasts are obtained using $p=2$, $q=1$; RR forecasts are obtained using $p=13$, $\lambda=494$. MSFEs are relative to the random walk forecasts of the cumulated target series. Bold figures indicate the best forecasting method. * (**) and [***] indicate significance at the 10% (5%) and [1%] levels for the test of equal MSFEs of the two best methods.

The second approach, labeled as GK, is the variant of SW proposed by Gröen and Kapetanios (2008), in which PLS is used in place of the PCs to extract the relevant factors from Z_t. In order to estimate the PLS factors of Z_t and the coefficients of Y_t^L, a switching algorithm is used. First, having fixed the coefficients of Y_t^L to an initial estimate, a conditional estimate of the PLS factors of Z_t is computed. Second, having fixed the PLS factors to their previously obtained estimates, a conditional estimate of the coefficients of Y_t^L is obtained. These two steps are iterated till numerical convergence occurs. Notice that PLS and GK differ only with respect to the treatment of the lags of y_t, which are not included in the factor structure according to the latter approach.

Table 7
EMP, relative MSFE.

Models	$h=1$	$h=3$	$h=6$	$h=12$
Sample: 1975–2007				
PLS	**0.583**	**0.503**	0.639	0.802
GK	0.623	0.527	**0.628**	**0.776**
PCR	0.627	1.097	1.454	1.709
SW	0.741	1.148	1.468	1.712
RR	0.642	1.013	1.434	1.717
Sample: 1975–1984				
PLS	**0.596**	**0.504***	**0.645**	**0.771**
GK	0.676	0.594	0.708	0.827
PCR	0.601	1.112	1.462	1.687
SW	0.684	1.154	1.468	1.686
RR	0.680	1.034	1.442	1.676
Sample: 1985–2007				
PLS	0.560	0.502	0.630	0.838
GK	**0.528**	**0.426****	**0.522*****	**0.721**
PCR	0.676	1.074	1.444	1.735
SW	0.843	1.138	1.468	1.742
RR	0.575	0.983	1.423	1.764

PLS forecasts are obtained using $p=2$, $q=4$; GK forecasts are obtained using $p=6$, $q=5$; PCR forecasts are obtained using $p=2$, $q=1$; SW forecasts are obtained using $p=2$, $q=1$; RR forecasts are obtained using $p=3$, $\lambda=57$. See the notes for Table 6 for further details.

Table 8
FYFF, relative MSFE.

Models	$h=1$	$h=3$	$h=6$	$h=12$
Sample: 1975–2007				
PLS	**0.767**	**0.899**	0.884	**0.891**
GK	0.803	0.986	**0.868**	1.022
PCR	0.874	0.936	0.920	0.912
SW	0.883	0.909	0.907	0.927
RR	1.181	0.962	1.020	0.941
Sample: 1975–1984				
PLS	**0.763**	**0.881**	0.845	**0.857**
GK	0.802	0.957	**0.793**	0.987
PCR	0.882	0.948	0.936	0.938
SW	0.892	0.912	0.921	0.967
RR	1.061	0.917	1.005	0.966
Sample: 1985–2007				
PLS	0.818	1.040	1.020	0.951
GK	0.810	1.215	1.133	1.084
PCR	**0.786**	**0.840*****	0.863	0.867
SW	0.786	0.890	**0.857**	**0.855****
RR	2.470	1.320	1.073	0.897

PLS forecasts are obtained using $p=1$, $q=7$; GW forecasts are obtained using $p=4$, $q=8$; PCR forecasts are obtained using $p=2$, $q=3$; SW forecasts are obtained using $p=2$, $q=2$; RR forecasts are obtained using $p=9$, $\lambda=85.5$. See the notes for Table 6 for further details.

Finally, for PLS, GK, PCR and SW the regression coefficients β^h are estimated by generalized least squares, allowing for both heteroskedasticity and autocorrelation of order (h-1).

Most of factor model literature suggests the use of information criteria for selecting the number of factors. However, there are no sound theoretical justifications for the use of information criteria for PLS and PCR, although these criteria are sometimes used by practitioners. Hence, we compare the results that we obtain for models selected with both information criteria and a cross-validation approach.

More in detail, the number of components q to be considered in PLS, PCR, SW and GK, the shrinking parameter λ for RR, as well as the number of lags p to be used in each method, are fixed either using information criteria on the sample 1959.01–1974.12 or by minimizing the 3-step ahead MSFE that is computed using the training sample 1959.01–1969.12 and the validation sample 1970.01–1974.12.[3] The maximum values for p and q are, respectively, 13 and 10. Finally, following De Mol et al. (2008), we choose the shrinking parameter among $\lambda/N=[0.5,1,2,5,10]$, where $N=19p$. The comparison between the two approaches indicates that cross-validation generally performs better than information criteria. This finding is in line with the analytical result by Hansen (2009), who proved that good in sample fit translates into poor out of sample fit. Hence, we document the results relative to the models selected by cross validation only.[4]

In order to check whether the differences between the MSFEs of the two best forecasting methods are statistically significant, we performed the version of the Diebold and Mariano (1995) test by Harvey et al. (1997). In particular the null hypothesis of equal MSFEs is tested against the alternative that the second best forecasting model has a larger MSFE.[5]

[3] Based on the MC results, in the cross-validation approach we use the 3-step ahead MSFE. Indeed, in simulations the most accurate PCR and PLS forecasts for $h=1$ are often obtained for models having a different number of components than the true one.

[4] The results relative to models selected by information criteria are available upon request.

[5] Giacomini and White (2006) provided an asymptotic framework that allows for using the Diebold and Mariano (1995) test with both nested models and the use of unconventional estimation methods.

Table 9
CPI, relative MSFE.

Models	$h=1$	$h=3$	$h=6$	$h=12$
Sample: 1975–2007				
PLS	0.900	0.902	**0.759**	**0.502**
GK	0.918	**0.890**	0.774	0.537
PCR	1.039	0.971	0.822	0.569
SW	**0.893**	0.913	0.799	0.518
RR	0.951	0.948	0.771	0.513
Sample: 1975–1984				
PLS	**0.804**	0.811	0.688	0.404
GK	0.828	**0.750**	**0.634**	0.405
PCR	0.979	0.929	0.778	0.504
SW	**0.804**	0.820	0.751	0.462
RR	0.812	0.772	0.660	0.410
Sample: 1985–2007				
PLS	0.975	**0.969**	**0.834**	0.626
GK	0.989	0.996	0.923	0.704
PCR	1.085	1.003	0.870	0.652
SW	**0.963**	0.984	0.851	**0.589***
RR	1.061	1.082	0.890	0.643

PLS forecasts are obtained using $p=4$, $q=2$; GW forecasts are obtained using $p=4$, $q=2$; PCR forecasts are obtained using $p=4$, $q=7$; SW forecasts are obtained using $p=8$, $q=1$; RR forecasts are obtained using $p=10$, $\lambda=950$. See the notes for Table 6 for further details.

Finally, in order to take into account the effects of the "Great Moderation", we consider three forecast evaluation samples: 1975.01–2007.12, 1975.01–1984.12 (pre-Great Moderation), and 1985.01–2007.12 (post-Great Moderation).

Tables 6–9 report the MSFEs relative to the naive random walk forecasts of the cumulated target series as well as the test results of equal MSFEs of the two best forecasting methods.

The empirical findings indicate that PLS delivers the most accurate forecasts in about 58.3% of the cases. It performs best for IPI in any sample, EMP in pre-Great Moderation period, FYFF in the full sample and in the pre-Great Moderation period. The second best performer is GK, which results slightly superior to SW. In particular, the former method provides the best forecasts of EMP in post-Great Moderation period whereas the latter is a serious contender to PLS for inflation forecasting. PCR and, especially, RR perform disappointingly compared with the simulation results.

Turning to the forecasting encompassing test results, we notice that just one fourth of the differences in MSFEs of the two best methods are significant. Again, PLS performs significantly better than its closest competitor in about 58.3% of the cases.

Looking in greater detail at the relative merits of the two best methods, PLS might be preferred to GK for both forecasting performances and computational reasons. Indeed, there are apparently no clear advantages in resorting to the rather involved iterative scheme suggested by Gröen and Kapetanios (2008).

5. Conclusions

In this paper we have examined the forecasting performances of various models in a medium-*N* environment. Moreover, we have argued that under the so-called Helland and Almoy condition (Helland, 1990; Helland and Almoy, 1994), both PCR and PLS provide estimates of a stable dynamic regression model that are consistent as *T* only diverges.

Our Monte Carlo results, obtained by simulating a 20-dimensional VAR(2) process that satisfy the Helland and Almoy condition, have revealed that PLS often outperforms the competitors, especially when the sample size *T* and the number of the relevant components become larger.

In the empirical application, we have forecasted, by a variety of competing models, four US monthly time series using similar variables as in the medium dimension VAR model by Banbura et al. (2010). Interestingly, PLS has revealed to perform better than other, more well-known, forecasting methods. Moreover, we emphasize that the suggested PLS approach is computationally less demanding than the switching algorithm proposed by Gröen and Kapetanios (2008).

Acknowledgments

Previous drafts of this paper were presented at seminars at the universities of Bologna and Maastricht, the 30th Annual International Symposium on Forecasting in San Diego, and the 65th European Meeting of the Econometric Society in Oslo. We wish to thank Alain Hecq and Alessandra Luati, as well as two anonymous referees, for helpful comments and suggestions. The usual disclaimers apply.

References

Almoy, T., 1996. A simulation study on comparison of prediction methods when only a few components are relevant. Computational Statistics and Data Analysis 21, 87–107.
Banbura, M., Giannone, D., Reichlin, L., 2010. Large Bayesian VARs. Journal of Applied Econometrics 25, 71–92.
Boivin, J., Ng, S., 2006. Are more data always better for factor analysis. Journal of Econometrics 132, 169–194.
Caggiano, G., Kapetanios, G., Labhard, V., 2011. Are more data always better for factor analysis? Results for the euro area, the six largest euro area countries and the UK. Journal of Forecasting 30, 736–752.
Centoni, M., Cubadda, G., Hecq, A., 2007. Common shocks, common dynamics, and the international business cycle. Economic Modelling 24, 149–166.
Chun, H., Keleş, S., 2010. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72, 3–25.
Cubadda, G., 2007. A unifying framework for analyzing common cyclical features in cointegrated time series. Computational Statistics and Data Analysis 52, 896–906.
Cubadda, G., Hecq, A., 2011. Testing for common autocorrelation in data-rich environments. Journal of Forecasting 30, 325–335.
Cubadda, G., Hecq, A., Palm, F.C., 2009. Studying co-movements in large multivariate models prior to multivariate modelling. Journal of Econometrics 148, 25–35.
De Mol, C., Giannone, D., Reichlin, L., 2008. Forecasting using a large number of predictors: is Bayesian regression a valid alternative to principal components? Journal of Econometrics 146, 30–66.
Diebold, F.X., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.
Eickmeier, S., Ng, T., 2011. Forecasting national activity using lots of international predictors: an application to New Zealand. International Journal of Forecasting 27, 496–511.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2003. The generalized factor model: consistency and rates. Journal of Econometrics 119, 231–255.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2005. The generalized dynamic factor model: one-sided estimation and forecasting. Journal of the American Statistical Association 100, 830–840.
Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74, 1545–1578.
Gröen, J.J., Kapetanios, G., 2008. Revisiting useful approaches to data-rich macroeconomic forecasting. Federal Reserve Bank of New York Staff Reports, p. 327.
Hadi, A.S., Ling, R.F., 1998. Some cautionary notes on the use of principal components regression. The American Statistician 52, 15–19.
Hansen, P.R. (2009), In sample fit and out-of-sample fit: their joint distribution and its implications for model selection, *mimeo*.
Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. International journal of forecasting 13, 281–291.
Helland, I.S., 1990. Partial least squares regression and statistical models. Scandinavian Journal of Statistics 17, 97–114.
Helland, I.S., Almoy, T., 1994. Comparison of prediction methods when only a few components are relevant. Journal of the American Statistical Association 89, 583–591.
Lin, J. and R.S. Tsay (2006), Comparisons of forecasting methods with many predictors. *Mimeo*.
Næs, T., Helland, I.S., 1993. Relevant components in regression. Scandinavian Journal of Statistics 20, 239–250.
Stock, J.H., Watson, M.W., 2002a. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.
Stock, J.H., Watson, M.W., 2002b. Macroeconomic forecasting using diffusion indexes. Journal of Business and Economics Statistics 20, 147–162.
Stoica, P., Söderström, T., 1998. Partial least squares: a first-order analysis. Scandinavian Journal of Statistics 25, 17–24.
Watson, M.W., 2003. In: Dewatripont, M., Hansen, L., Turnovsky, S. (Eds.), Macroeconomic Forecasting Using Many Predictors: Advances in Econometrics, Theory and Applications, Eighth World Congress of the Econometric Society, Vol. III, pp. 87–115.
Wold, H., 1985. Partial least squares. In: Kotz, S., Johnson, N.L. (Eds.), Encyclopedia of the Statistical Sciences, vol. 6. John Wiley & Sons, pp. 581–591.