A Medium-N Approach to Macroeconomic Forecasting

Gianluca Cubadda*
Università di Roma "Tor Vergata", Roma, Italy

Barbara Guardabascio
Università di Roma "Tor Vergata", Roma, Italy

January 2010

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. This is achieved by examining the conditions under which partial least squares and principal component regression provide consistent estimates of a stable autoregressive distributed lag model as only the number of observations, T, diverges. We show both by simulations and empirical applications that the proposed methods compare well to models that are widely used in macroeconomic forecasting.

Keywords: Macroeconomic forecasting; partial least squares; principal component regression.

*Corresponding author: Dipartimento SEFEMEQ, Università di Roma "Tor Vergata", Via Columbia 2, 00133 Roma, Italy. Tel. +39 06 72595847, Fax +39 06 2040219, email gianluca.cubadda@uniroma2.it.

1 Introduction

Growing attention has recently been devoted to forecasting economic time series in a data rich framework. 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 components [PCs] 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, Heij *et al.* (2007) and Groen and Kapetanios (2008) have, respectively, proposed principal covariate regression and 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, and Banbura *et al.* (2010) also found that a vector autoregressive [VAR] model with 20 key macroeconomic indicators forecasts as well as a larger model of 131 variables.

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. Particularly, we show that, under the so-called Helland & 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 autoregressive distributed lag [ADL] 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 per-

formances 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 & 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 ADL models estimated by PCR and, especially, PLS forecast well when compared to both OLS and RR.

In the empirical application, we forecast four US macro time series by a rich variety of methods using the same variables as in the 20-dimensional VAR model in Banbura *et al.* (2010). The empirical findings indicate that forecasting methods based on PLS outperform the competitors. Interestingly, Groen and Kapetanios (2008) reached a similar conclusion 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 an empirical applications to US economic variables. Finally, Section 5 concludes.

2 Theory

Let us suppose that the stationary and ergodic scalar time series to be forecasted, y_{t+1}, is generated by the following regression model

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

where X_t is N-vector of stationary and ergodic time series, possibly including lags of y_t, ε_t is i.i.d. with $\mathrm{E}(\varepsilon_t) = 0$, $\mathrm{E}(\varepsilon_t^2) = \sigma_\varepsilon^2$, $\mathrm{E}(\varepsilon_t^4) < \infty$, and such that $\mathrm{E}(\varepsilon_t | X_t) = 0$. Moreover, we assume that deterministic elements have preliminarily been removed from both time series y_t and X_t, and that each element of X_t has been standardized to 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 *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 *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 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, matrices of dimension $N \times q$ and $N \times (N-q)$ such that $\theta'\theta = I_q$, $\theta'_\perp \theta_\perp = I_{N-q}$, $\theta\theta' = I_N - \theta_\perp \theta'_\perp$, $\mathrm{E}(R_t E'_t) = 0$, and $\Sigma_{xy} = \theta \mathrm{E}(R_t y_{t+1})$. 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.

Whether condition (2) is generally appropriate for macroeconomic time series is an empirical issue that we will consider later in Section 4.

Notice that condition (2) 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:

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

where $F_t = \Lambda_q^{-1} \Upsilon'_q X_t$. Hence, since $y_{t+1|t} = \mathrm{E}(y_{t+1}|X_t)$ 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.

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 equation (3), it is clear under Condition (2) we have that $\beta_{PCR} = \beta$.

However, notice that condition (2) does not require that the eigenvalues associated to the eigenvectors Υ_q are the q largest ones. In empirical applications of PCR, it is clear the one has to select the relevant principal components and the sample eigenvalues offer no guidance on this choice. As shown by Helland (1990), PLS offer an effective way to overcome this problem. Indeed, let $\beta'_{PLS} X_t$ indicate the PLS prediction of y_{t+1}, where

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

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

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

with $\omega_1 = \Sigma_{xy}$. It follows by induction from (6) that Ω_q lies in the space spanned by

$$(\Sigma_{xy}, \Sigma_{xx}\Sigma_{xy}, ..., \Sigma_{xx}^{q-1}\Sigma_{xy}),$$

and hence, under condition (2), in the space spanned by the eigenvectors Υ_q. Finally, this last result implies that $\omega_{q+1} = 0$ 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

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

for $i = 1, .., q$, where

$$\begin{aligned} f_{i,t} &= \omega_i' V_{i-1,t}, \\ \omega_i &= \mathrm{E}(V_{i-1,t} y_{t+1}), \\ \phi_i &= \mathrm{E}(f_{i,t} V_{i-1,t}) / \mathrm{E}(f_{i,t}' f_{i,t}) \end{aligned}$$

It is easy to see from equation (7) that the PLS factors $f_{i,t}$ are uncorrelated with one other and that they are a non-singular linear transformation of $\Omega_q' X_t$. Hence, $\beta_{PLS}' X_t$ may be equivalently obtained by a linear regression of y_{t+1} on $(f_{1,t}, ...f_{q,t})'$.

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 principal components, 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 (2) the sample versions of (4) and (5) are consistent estimators of β as $T \to \infty$ by the Slutsky's theorem. Helland and Almoy (1994) compared PCR and PLS on the basis of their expected prediction error and concluded that no method asymptotically dominates the other.

In the next sections we will assess the forecasting performances of PCR and PLS both by simulations and empirical examples.

3 Monte Carlo analysis

Apart from consistency, not much is known on the statistical properties of PLS and PCR. Hence, in this Section we carry out a Monte Carlo study to evaluate the forecasting performances of these methods in a medium-N framework.

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 q 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 q 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 U_t

$$Y_t = Q H_t,$$

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. $\mathrm{N}_n(0, I_n)$. Hence, series Y_t follow a stationary VAR(2) with a reduced-rank structure (see,

inter alia, Cubadda (2007) and Cubadda *et al.* (2009) for the statistical and economic implications of this kind of structures). It follows that each element of Y_t is generated by a stable ADL model with the same form as (1), where y_t is a generic element of the vector series Y_t, ε_t is the corresponding element of $Q\epsilon_t$, and $X_t = [Y_t', Y_{t-1}']'$.

We notice that the relevant and irrelevant components of X_t are respectively given by

$$R_t = [Y_{t-1}' Q_{\cdot q}, Y_{t-2}' Q_{\cdot q}]'$$

and

$$E_t = [Y_{t-1}' Q_{\cdot n-q}, Y_{t-2}' Q_{\cdot n-q}]',$$

where $[Q_{\cdot q}, Q_{\cdot n-q}] = Q$, and $Q_{\cdot q}$ is an $n \times q-$matrix. Hence, condition (2) is satisfied.

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

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 $X_\tau' \widehat{\beta}^h_\lambda$ where

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

and λ is a shrinkage scalar parameter. Since De Mol *et al.* (2008) document 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 $X_\tau' \widehat{\beta}^h_{PCR}$ where

$$\widehat{\beta}^h_{PCR} = \widehat{\Upsilon}_q \widehat{\Lambda}_q^{-1} \widehat{\Upsilon}_q' X'y,$$

and $X\widehat{\Upsilon}_q$ 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 $X_\tau' \widehat{\beta}^h_{PLS}$ where

$$\widehat{\beta}^h_{PLS} = (\widehat{F}'\widehat{F})^{-1}\widehat{F}'y,$$

$\widehat{F} = (\widehat{F}_1, ..., \widehat{F}_{T-h})'$, and $\widehat{F}_t = (\widehat{f}_{1,t}, ...\widehat{f}_{q,t})'$ is obtained recursively from equation (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 an AR(2) forecast. To construct these relative MSFEs, we simulate systems of $n = 20$ variables (i.e. $N = 40$ predictors) with $q = 2, 4, 6, 8$. We generate $T + 170$ observations of the vector series Y_t for $T = 240, 360, 480$ corresponding to $20, 30, 40$ years of monthly observations. 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 for $h = 1, 3, 6, 12$, and the intermediate T observations are used to estimate the various models. The results, reported in Tables 1-3, are based on 5000 replications of series y_t.

TABLE 1
Simulations, Relative MSFE

$T = 240$					
$q = 2$					
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$	mean
PLS	1.000	0.944	0.966	0.978	0.972
PCR	0.948	0.954	0.967	0.978	0.962
OLS	1.051	1.100	1.136	1.169	1.114
RR(0.5)	1.010	1.057	1.090	1.118	1.069
RR(1)	0.986	1.031	1.062	1.088	1.041
RR(2)	0.959	1.001	1.029	1.052	1.010
RR(5)	0.932	0.968	0.991	1.010	0.975
RR(10)	0.930	0.959	0.978	0.994	0.961
$q = 4$					
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$	mean
PLS	0.938	0.913	0.951	0.978	0.945
PCR	0.902	0.930	0.954	0.980	0.941
OLS	0.948	1.045	1.105	1.160	1.064
RR(0.5)	0.909	0.999	1.053	1.102	1.016
RR(1)	0.890	0.975	1.026	1.079	0.990
RR(2)	0.871	0.950	0.996	1.037	0.964
RR(5)	0.866	0.930	0.991	0.967	0.941
RR(10)	0.891	0.939	0.962	0.991	0.944
$q = 6$					
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$	mean
PLS	0.862	0.887	0.937	0.978	0.918
PCR	0.859	0.912	0.947	0.981	0.924
OLS	0.863	0.995	1.080	1.150	1.022
RR(0.5)	0.823	0.949	1.024	1.087	0.972
RR(1)	0.812	0.928	0.998	1.056	0.948
RR(2)	0.803	0.908	0.971	1.023	0.926
RR(5)	0.815	0.898	0.949	0.991	0.913
RR(10)	0.855	0.915	0.954	0.985	0.925
$q = 8$					
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$	mean
PLS	0.774	0.853	0.923	0.969	0.880
PCR	0.800	0.881	0.933	0.972	0.896
OLS	0.773	0.931	1.038	1.129	0.968
RR(0.5)	0.744	0.889	0.984	1.064	0.920
RR(1)	0.738	0.873	0.961	1.031	0.901
RR(2)	0.740	0.861	0.940	1.001	0.886
RR(5)	0.773	0.866	0.930	0.975	0.886
RR(10)	0.834	0.897	0.946	0.976	0.911

Notes: MSFE are relative to an AR(2) forecast. RR(λ/N) indicates RR with a shrinkink parameter λ.

TABLE 2
Simulations, Relative MSFE

$T = 360$					
$q = 2$					
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$	mean
PLS	0.970	0.949	0.971	0.986	0.969
PCR	0.940	0.958	0.971	0.986	0.964
OLS	0.991	1.038	1.069	1.105	1.051
RR(0.5)	0.975	1.021	1.050	1.083	1.032
RR(1)	0.964	1.009	1.037	1.068	1.019
RR(2)	0.949	0.993	1.019	1.048	1.002
RR(5)	0.930	0.970	0.993	1.018	0.978
RR(10)	0.926	0.960	0.980	1.000	0.963
$q = 4$					
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$	mean
PLS	0.874	0.894	0.936	0.960	0.916
PCR	0.871	0.913	0.940	0.961	0.921
OLS	0.874	0.959	1.018	1.066	0.979
RR(0.5)	0.860	0.941	0.999	1.043	0.961
RR(1)	0.852	0.931	0.987	1.030	0.950
RR(2)	0.843	0.920	0.972	1.011	0.936
RR(5)	0.841	0.909	0.954	0.987	0.923
RR(10)	0.858	0.914	0.952	0.978	0.921
$q = 6$					
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$	mean
PLS	0.803	0.855	0.922	0.963	0.888
PCR	0.818	0.890	0.930	0.964	0.901
OLS	0.800	0.918	0.997	1.062	0.945
RR(0.5)	0.786	0.899	0.975	1.035	0.924
RR(1)	0.780	0.889	0.962	1.020	0.913
RR(2)	0.775	0.879	0.946	1.000	0.976
RR(5)	0.783	0.872	0.930	0.976	0.890
RR(10)	0.809	0.881	0.929	0.967	0.895
$q = 8$					
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$	mean
PLS	0.731	0.837	0.905	0.950	0.856
PCR	0.769	0.863	0.911	0.951	0.874
OLS	0.730	0.872	0.963	1.038	0.901
RR(0.5)	0.719	0.854	0.940	1.009	0.881
RR(1)	0.717	0.847	0.929	0.995	0.872
RR(2)	0.721	0.842	0.918	0.978	0.864
RR(5)	0.745	0.846	0.910	0.960	0.865
RR(10)	0.791	0.869	0.920	0.960	0.883

Note: See the notes for Table 1.

TABLE 3
Simulations, Relative MSFE

$T = 480$					
$q = 2$					
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$	mean
PLS	0.944	0.939	0.963	0.973	0.955
PCR	0.926	0.945	0.960	0.973	0.951
OLS	0.956	1.001	1.029	1.058	1.011
RR(0.5)	0.947	0.991	1.019	1.047	1.001
RR(1)	0.941	0.984	1.011	1.038	0.994
RR(2)	0.932	0.974	1.000	1.025	0.983
RR(5)	0.919	0.958	0.982	1.004	0.966
RR(10)	0.914	0.949	0.971	0.989	0.952
$q = 4$					
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$	mean
PLS	0.869	0.901	0.942	0.966	0.920
PCR	0.873	0.922	0.947	0.959	0.928
OLS	0.867	0.949	1.001	1.044	0.965
RR(0.5)	0.858	0.938	0.990	1.031	0.954
RR(1)	0.852	0.931	0.987	1.030	0.950
RR(2)	0.846	0.923	0.971	1.009	0.937
RR(5)	0.842	0.912	0.946	0.989	0.925
RR(10)	0.849	0.910	0.948	0.977	0.919
$q = 6$					
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$	mean
PLS	0.766	0.851	0.913	0.952	0.871
PCR	0.800	0.879	0.922	0.955	0.889
OLS	0.762	0.874	0.958	1.020	0.904
RR(0.5)	0.756	0.865	0.955	1.004	0.892
RR(1)	0.753	0.860	0.938	0.995	0.886
RR(2)	0.753	0.855	0.929	0.983	0.880
RR(5)	0.762	0.854	0.918	0.965	0.875
RR(10)	0.785	0.863	0.917	0.957	0.879
$q = 8$					
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$	mean
PLS	0.708	0.836	0.906	0.951	0.850
PCR	0.753	0.865	0.917	0.954	0.872
OLS	0.706	0.845	0.943	1.012	0.878
RR(0.5)	0.701	0.848	0.931	0.997	0.867
RR(1)	0.700	0.837	0.924	0.987	0.862
RR(2)	0.702	0.833	0.916	0.975	0.857
RR(5)	0.721	0.836	0.901	0.960	0.856
RR(10)	0.756	0.852	0.911	0.956	0.867

Note: See the notes for Table 1.

The results indicate that OLS is generally outperformed by the competitors. Indeed, OLS performs similarly as the other methods only for $T = 480$ and $q = 8$ and worse in all the other cases. This finding suggests that the cost of ignoring restrictions on β given by (2) is high in a medium-N framework even when the sample size is large. PLS performs better as both T and q become larger. In particular, PLS always provides the most accurate forecast when $q = 8$. In contrast, PCR often forecasts best when q is small. The performance of RR depends crucially on the choice of the shrinking parameter λ. In general, the larger is q, the smaller should be λ. 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 (2) is satisfied. In the next Section, we evaluate the relative merits of PLS in a empirical exercise.

4 Empirical application

In order to perform our empirical out-of-sample forecasting exercise, we use the same data-set as Banbura *et al.* (2010) for their medium dimension VAR model. It consists of 20 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 time span is from 1959.01 through to 2003.12. We apply logarithms to most of the series with the exception of those that are already expressed in rates. In order to render all variables stationary, the same transformations as in Banbura *et al.* (2010) are applied, thus obtaining the vector series Y_t. 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 target series is

$$y_{t+h}^h = \frac{(1-L^h)}{(1-L)} y_{t+h}, \quad , h = 1, 3, 6, 12$$

where L is the usual lag operator, 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, labelled as SW, is the Stock and Watson (2002a, 2002b) dynamic factor model, which computes the h-step ahead forecast of $y_{\tau+h}^h$ as $W_\tau' \widehat{\beta}_{SW}^h$, where $W_t = [Z_\tau' \widehat{\Psi}_q, Y_t^{L\prime}]'$, $Y_t = [y_t, Z_t']'$, $Y_t^L = [y_t, .., y_{t-p+1}]'$, $\widehat{\Psi}_q$ are the eigenvectors associated with the q largest eigenvalues of $Z'Z$, and $Z = [Z_1, ..., Z_T]'$

The second approach, labelled as GK, is the variant of SW proposed by Groen 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.

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 $(h-1)$.

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 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. 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 $N = 20p$.

The following tables report the MSFE relative to the naive random walk forecast for all the models considered. In order to take into account the Great Moderation effects, we consider three forecast evaluation samples: 1975.01-2003.12, 1975.01-1984.12 (pre-Great Moderation), 1985.01-2003.12 (post-Great Moderation).

The empirical findings suggest that, in each of the three evaluation samples that we consider, PLS and GK perform similarly and outperform the competitors in most cases. Looking in greater detail at the relative merits of the best performers, PLS (GK) forecasts better IP and FYFF (EMP), whereas they are almost equivalent for CPI in the largest evaluation sample. On the basis of the similar forecasting performances, PLS might be preferred for computational reasons. Indeed, there are apparently no clear advantages in resorting to the rather involved iterative scheme suggested by Groen and Kapetanios (2008).

Finally, we notice that it is not always the case that the forecasting performances worse during the Great Moderation. Indeed, whereas IPI and CPI exhibit lower MSFEs in the period 1975-1984, we reach the opposite conclusion for EMP and the results are mixed for FYFF, depending on the considered forecast horizon.

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 & Almoy condition (Helland, 1990; Helland and Almoy, 1994), both PCR and PLS provide estimates of a stable ADL 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 & Almoy condition, have revealed that PLS often outperforms the competitors, especially when the sample size T and the number of the relevant components q become larger.

In the empirical application, we have forecasted, by a variety of competing models, four US monthly time series using the same variables as in the 20-

TABLE 4
IPI, Relative MSFE

Sample: 1975-2003				
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
PLS	0.821	0.782	0.879	0.899
GK	0.857	0.846	0.963	1.051
PCR	0.828	1.034	1.187	1.357
SW	0.816	1.050	1.184	1.351
RR	0.976	1.013	1.178	1.373
Sample: 1975-1984				
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
PLS	0.737	0.726	0.834	0.859
GK	0.788	0.803	0.965	1.037
PCR	0.723	0.982	1.131	1.271
SW	0.788	1.033	1.146	1.260
RR	0.897	0.934	1.096	1.269
Sample: 1985-2003				
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
PLS	0.943	0.881	0.945	0.951
GK	0.959	0.918	0.954	1.063
PCR	0.946	1.119	1.274	1.467
SW	0.854	1.079	1.244	1.468
RR	1.092	1.151	1.302	1.506

Note: MSFE are relative to a Random Walk forecast. PLS forecasts are obtained using $p = 12$, $q = 2$; GK forecasts are obtained using $p = 8$, $q = 3$; PCR forecasts are obtained using $p = 1$, $q = 2$; SW forecasts are obtained using $p = 2$, $q = 2$; RR forecasts are obtained using $p = 1$, $\lambda = 40$.

TABLE 5
EMP, Relative MSFE

Sample: 1975 - 2003				
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
PLS	0.580	0.521	0.670	0.815
GK	0.603	0.516	0.611	0.742
PCR	0.643	1.103	1.437	1.666
SW	0.684	1.099	1.428	1.667
RR	0.656	0.996	1.406	1.672
Sample: 1975 -1984				
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
PLS	0.601	0.556	0.723	0.850
GK	0.649	0.594	0.709	0.825
PCR	0.651	1.138	1.496	1.771
SW	0.645	1.118	1.474	1.761
RR	0.671	1.035	1.473	1.768
Sample: 1985 - 2003				
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
PLS	0.539	0.460	0.591	0.771
GK	0.510	0.389	0.470	0.643
PCR	0.628	1.040	1.347	1.530
SW	0.761	1.065	1.357	1.545
RR	0.626	0.929	1.308	1.545

Note: MSFE are relative to a Random Walk forecast. PLS forecasts are obtained using $p = 2$, $q = 4$;GK forecasts are obtained using $p = 6$, $q = 4$; PCR forecasts are obtained using $p = 2$, $q = 1$; SW forecasts are obtained using $p = 2$, $q = 4$; RR forecasts are obtained using $p = 3$, $\lambda = 60$.

TABLE 6
FYFF, Relative MSFE

Sample: 1975 - 2003				
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
PLS	0.760	0.813	0.926	0.898
GK	0.935	1.287	1.043	0.926
PCR	0.867	0.961	0.946	0.935
SW	0.890	0.908	0.912	0.951
RR	1.197	0.960	1.036	0.982
Sample: 1975 -1984				
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
PLS	0.758	0.828	0.971	0.929
GK	0.925	1.314	1.066	0.969
PCR	0.875	0.971	0.959	0.947
SW	0.885	0.907	0.917	0.972
RR	1.051	0.901	1.018	0.974
Sample: 1985 - 2003				
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
PLS	0.782	0.681	0.713	0.820
GK	1.051	1.021	0.937	0.823
PCR	0.764	0.873	0.888	0.906
SW	0.931	0.924	0.887	0.899
RR	2.878	1.500	1.118	1.003

Note: MSFE are relative to a Random Walk forecast. PLS forecasts are obtained using $p = 2$, $q = 3$; GW forecasts are obtained using $p = 11$, $q = 2$; PCR forecasts are obtained using $p = 2$, $q = 2$; SW forecasts are obtained using $p = 2$, $q = 5$; RR forecasts are obtained using $p = 9$, $\lambda = 90$.

TABLE 7
CPI, Relative MSFE

Sample: 1975 - 2003				
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
PLS	0.831	0.828	0.663	0.420
GK	0.862	0.804	0.868	0.385
PCR	0.884	0.872	0.711	0.470
SW	0.767	0.808	0.698	0.430
RR	0.958	0.900	0.705	0.452
Sample: 1975 -1984				
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
PLS	0.768	0.787	0.628	0.401
GK	0.835	0.726	0.578	0.358
PCR	0.829	0.864	0.712	0.483
SW	0.721	0.745	0.670	0.433
RR	0.791	0.756	0.603	0.385
Sample: 1985 - 2003				
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
PLS	0.909	0.897	0.722	0.455
GK	0.895	0.935	0.867	0.436
PCR	0.953	0.884	0.707	0.445
SW	0.824	0.910	0.742	0.425
RR	1.175	1.135	0.868	0.571

Note: MSFE are relative to a Random Walk forecast. PLS forecasts are obtained using $p = 4$, $q = 2$; GW forecasts are obtained using $p = 9$, $q = 2$; PCR forecasts are obtained using $p = 10$, $q = 1$; SW forecasts are obtained using $p = 12$, $q = 1$; RR forecasts are obtained using $p = 10$, $\lambda = 400$.

dimensional VAR model in Banbura *et al* (2010). Interestingly, PLS has revealed to perform similarly as the procedure proposed by Groen and Kapetanios (2008) and better than other, more well-known, forecasting methods. However, we emphasize the our PLS approach is computationally less demanding than the GK switching algorithm.

References

[1] 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

[2] Banbura, M., Giannone, D., and L. Reichlin (2010), Large Bayesian VARs, *Journal of Applied Econometrics*, 25, 71 - 92

[3] Boivin, J. and S. Ng (2006), Are more data always better for factor analysis, *Journal of Econometrics*, 132, 169–194.

[4] Cubadda, G. (2007), A unifying framework for analyzing common cyclical features in cointegrated time series, *Computational Statistics and Data Analysis,* 52, 896-906

[5] Cubadda, G., Hecq, A. and F.C. Palm (2009), Studying comovements in large multivariate models prior to multivariate modelling, *Journal of Econometrics*, 148, 25-35.

[6] De Mol, C., Giannone, D., and L. Reichlin (2008), Forecasting using a large number of predictors: is Bayesian regression a valid alternative to principal components?, *Journal of Econometrics*, 30-66.

[7] Forni, M., Hallin, M., Lippi, M., and L. Reichlin (2003), The Generalized Factor Model: Consistency and Rates, *Journal of Econometrics*, 119, 231 - 255.

[8] Forni, M., Lippi, M., Hallin, M., and L. Reichlin. (2005), The generalized dynamic factor model: One-sided estimation and forecasting, *Journal of the American Statistical Association*, 100, 830- 840.

[9] Gröen, J.J., and G. Kapetanios (2008), Revisiting useful approaches to data-rich macroeconomic forecasting, *Federal Reserve Bank of New York Staff Reports*, 327.

[10] Helland, I.S. (1990), Partial least squares regression and statistical models, *Scandinavian Journal of Statistics*, 17, 97-114.

[11] Helland, I.S. and T. Almoy (1994), Comparison of prediction methods when only a few components are relevant, *Journal of the American Statistical Association*, 89, 583-591.

[12] Heij, C., Groenen, P., and D. van Dijk (2007), Forecast comparison of principal component regression and principal covariate regression, *Computational Statistics & Data Analysis*, 51, 3612-3625.

[13] Næs, T. and I.S. Helland (1993), Relevant components in regression, *Scandinavian Journal of Statistics*, 20, 239-250.

[14] Stock J.H. and M.W. Watson (2002a), Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association*, 97, 1167-1179.

[15] Stock J.H. and M.W. Watson (2002b), Macroeconomic forecasting using diffusion indexes, *Journal of Business and Economics Statistics*, 20, 147-162.

[16] Watson, M.W. (2003), Macroeconomic forecasting using many predictors, in Dewatripont, M., L. Hansen and S. Turnovsky (eds.), *Advances in Econometrics, Theory and Applications*, Eighth World Congress of the EconometricSociety, Vol. III, 87-115.30

[17] Wold, H. (1985), Partial least squares, in S. Kotz and N.L. Johnson (eds), *Encyclopedia of the Statistical Sciences*, John Wiley & Sons, 6, 581–591.