Journal of Econometrics 186 (2015) 294–316

Contents lists available at ScienceDirect

Journal of Econometrics

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

The three-pass regression filter: A new approach to forecasting using many predictors[☆]

Bryan Kelly[a,*], Seth Pruitt[b,c]

[a] University of Chicago, Booth School of Business, 5807 S. Woodlawn Ave., Chicago, IL 60637, United States
[b] Board of Governors of the Federal Reserve System, Washington DC, United States
[c] Arizona State University, W.P. Carey School of Business, Tempe, AZ 852877, United States

ARTICLE INFO

Article history:
Available online 17 March 2015

JEL classification:
C22
C23
C53

Keywords:
Forecast
Factor model
Principal components
Constrained least squares
Partial least squares

ABSTRACT

We forecast a single time series using many predictor variables with a new estimator called the three-pass regression filter (3PRF). It is calculated in closed form and conveniently represented as a set of ordinary least squares regressions. 3PRF forecasts are consistent for the infeasible best forecast when both the time dimension and cross section dimension become large. This requires specifying only the number of relevant factors driving the forecast target, regardless of the total number of common factors driving the cross section of predictors. The 3PRF is a constrained least squares estimator and reduces to partial least squares as a special case. Simulation evidence confirms the 3PRF's forecasting performance relative to alternatives. We explore two empirical applications: Forecasting macroeconomic aggregates with a large panel of economic indices, and forecasting stock market returns with price–dividend ratios of stock portfolios.

1. Introduction

A common interest among economists and policymakers is harnessing vast predictive information to forecast important economic aggregates like national product or stock market value. However, it can be difficult to use this wealth of information in practice. If the predictors number near or more than the number of observations, the standard ordinary least squares (OLS) forecaster is known to be poorly behaved or nonexistent.[1]

How then does one effectively use vast predictive information? A solution well known in the economics literature views the data as generated from a model in which latent factors drive the systematic variation of both the forecast target, $\boldsymbol{y}$, and the matrix of predictors, $\boldsymbol{X}$. In this setting, the best prediction of $\boldsymbol{y}$ is infeasible since the factors are unobserved. As a result, a factor estimation step is required. The literature's benchmark method extracts factors that are significant drivers of variation in $\boldsymbol{X}$ and then uses these to forecast $\boldsymbol{y}$.

Our procedure springs from the idea that the factors that are *relevant* to $\boldsymbol{y}$ may be a strict subset of all the factors driving $\boldsymbol{X}$. Our method, called the three-pass regression filter (3PRF), selectively identifies only the subset of factors that influence the forecast target while discarding factors that are irrelevant for the target but that may be pervasive among predictors. The 3PRF has the advantage of being expressed in closed form and virtually instantaneous to compute.

This paper makes four main contributions. The first is to develop asymptotic theory for the 3PRF. We begin by proving that the estimator converges in probability to the infeasible best forecast in the (simultaneous) limit as cross section size N and time series dimension T become large. This is true even when variation in predictors is dominated by target-irrelevant factors. We then derive the limiting distributions for the estimated forecasts and predictive coefficients, and provide consistent estimators of asymptotic covariance matrices that can be used to perform inference. The second contribution of the paper is to verify the

☆ We thank Andrew Ang, Gray Calhoun, John Cochrane, Drew Creal, Rob Engle, Jim Hamilton, Jim Nason, Serena Ng, Jim Stock, Eric Swanson, Matt Taddy and Mark Watson and participants at University of Chicago (Booth), NBER Summer Institute EFWW 2011, Federal Reserve Board, Econometric Society N.A. Summer Meeting 2011, SoFiE Annual Meeting 2011, NBER-NSF Time Series 2011 Conference and SITE 2011 for helpful comments. We especially thank the co-editor and two anonymous referees for comments which improved the paper. We thank Mark Watson and Domenico Giannone for kindly sharing data and code.

* Corresponding author. Tel.: +1 773 702 8359.
E-mail addresses: bryan.kelly@chicagobooth.edu (B. Kelly), seth.pruitt@asu.edu (S. Pruitt).

[1] See Huber (1973) on the asymptotic difficulties of least squares when the number of regressors is large relative to the number of data points.

http://dx.doi.org/10.1016/j.jeconom.2015.02.011
0304-4076/

finite sample accuracy of our asymptotic theory through Monte Carlo simulations.

We also show that the method of partial least squares (PLS) is a special case of the 3PRF. Like partial least squares, the 3PRF can use the forecast target to discipline its dimension reduction. This emphasizes the covariance between predictors and target in the factor estimation step. But unlike PLS, the 3PRF also allows the econometrician to select additional disciplining variables, or factor proxies, on the basis of economic theory. Furthermore, because it is a special case of our methodology, the asymptotic theory we develop for the 3PRF applies directly to partial least squares. Recently Groen and Kapetanios (2009) showed the consistency of PLS under sequential N, T limits, while our approach proves consistency in the less restrictive simultaneous N, T limit. Those authors do not derive limiting distributions as we do here and so, to the best of our knowledge, our joint N and T asymptotics are new results to the PLS literature.

In our third contribution, we compare the 3PRF to other methods in order to illustrate the source of its improvement in forecasting performance. The economics literature has relied mainly on principal component regression (PCR) for forecasting problems involving many predictors, exemplified by Stock and Watson (1998, 2002a,b, 2006, 2012), Forni and Reichlin (1996, 1998), Bai and Ng (2002, 2006, 2008), Bai (2003) and Boivin and Ng (2006), among others.[2] Like the 3PRF, PCR can be calculated instantaneously for virtually any N and T. Stock and Watson's key insight is to condense information from the large cross section into a small number of predictive indices *before* estimating a linear forecast. PCR condenses the cross section according to *covariance within the predictors*. This identifies the factors driving the panel of predictors, some of which may be irrelevant for the dynamics of the forecast target, and uses those factors to forecast.

In contrast, the 3PRF condenses the cross section according to *covariance with the forecast target*. PCR must estimate *all* common factors among predictors to achieve consistency, including those that are irrelevant for forecasting. The 3PRF need only estimate the relevant factors, which are always less than or equal to the total number of factors required by PCR. While this difference is innocuous in large samples, it can be a crucial consideration in small samples.

We are not the first to investigate potential improvements upon PCR factor-based forecasts. Doz et al. (2012) propose quasi-maximum likelihood factor estimation as an alternative to PCR. Bai and Ng (2008) propose statistical thresholding rules that drop variables found to contain irrelevant information, building on the insights in Boivin and Ng (2006). In a similar vein, De Mol et al. (2008) propose Bayesian shrinkage methods. Thresholding and shrinkage methods are especially useful when relevant information is non-pervasive and confined to a subset of predictors. This does not solve the problem of pervasive irrelevant information among predictors. Our approach explicitly allows for both relevant and irrelevant pervasive factors.[3]

The final contribution of the paper is to provide empirical support for the 3PRF's strong forecasting performance in simulations and two separate empirical applications. We compare 3PRF to PCR, thresholding methods of Bai and Ng (2008), shrinkage methods of De Mol et al. (2008), and the factor analytic approach of Doz et al. (2012). Simulations show that the 3PRF often outperforms alternatives across a variety of factor model specifications. In empirical applications, we find that the 3PRF is a successful predictor of macroeconomic aggregates and equity market returns, and typically outperforms alternative methods.

The paper is structured as follows. Section 2 defines the 3PRF and proves its asymptotic properties. Section 3 reinterprets the 3PRF as a constrained least squares solution, then compares and contrasts it with partial least squares. Section 4 explores the finite sample performance of the 3PRF and other methods in Monte Carlo experiments. Section 5 reports empirical results for 3PRF and other methods' forecasts in asset pricing and macroeconomic applications. All proofs and supporting details are placed in the Appendix.

2. The three-pass regression filter

2.1. The estimator

There are several equivalent approaches to formulating our procedure, each emphasizing a related interpretation of the estimator. We begin with what we believe to be the most intuitive formulation of the filter, which is the sequence of OLS regressions that gives the estimator its name.

First we establish the environment wherein we use the 3PRF. There is a *target* variable which we wish to forecast. There exist many *predictors* which may contain information useful for predicting the target variable. The number of predictors N may be large and number near or more than the available time series observations T, which makes OLS problematic. Therefore we look to reduce the dimension of predictive information, and to do so we assume the data can be described by an approximate factor model. In order to make forecasts, the 3PRF uses *proxies*: These are variables, driven by the factors (and as we emphasize below, driven by *target-relevant* factors in particular), which we show are always available from the target and predictors themselves, but may alternatively be supplied to the econometrician on the basis of economic theory. The target is a linear function of a subset of the latent factors plus some unforecastable noise. The optimal forecast therefore comes from a regression on the true underlying relevant factors. However, since these factors are unobservable, we call this the *infeasible best forecast*.

We write $\boldsymbol{y}$ for the $T \times 1$ vector of the target variable time series from $2, 3, \ldots, T+1$.[4] Let $\boldsymbol{X}$ be the $T \times N$ matrix of predictors, $\boldsymbol{X} = (\boldsymbol{x}'_1, \boldsymbol{x}'_2, \ldots, \boldsymbol{x}'_T)' = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ that have been standardized to have unit time series variance. Note that we are using two different typefaces to denote the N-dimensional cross section of predictors observed at time t ($\boldsymbol{x}_t$), and the T-dimensional time series of the ith predictor ($\mathbf{x}_i$). This is to distinguish the time series of predictors from the cross section of predictors in Table 1. We denote the $T \times L$ matrix of proxies as $\boldsymbol{Z}$, which stacks period-by-period proxy data as $\boldsymbol{Z} = (\boldsymbol{z}'_1, \boldsymbol{z}'_2, \ldots, \boldsymbol{z}'_T)'$. We make no assumption on the relationship between N and T but assume $L \ll \min(N, T)$ in the spirit of dimension reduction. We provide additional details regarding the data generating processes for $\boldsymbol{y}$, $\boldsymbol{X}$ and $\boldsymbol{Z}$ in Assumption 1.

With this notation in mind, the 3PRF's regression-based construction is defined in Table 1. The first pass runs N separate *time series* regressions, one for each predictor. In these first pass regressions, the predictor is the dependent variable, the proxies are the regressors, and the estimated coefficients describe the sensitivity of the predictor to factors represented by the proxies. As we show later, proxies need not represent specific factors and

[2] The model investigated by Forni et al. (2000, 2004, 2005) concentrates on a frequency domain approach.

[3] We also demonstrate that the performance of 3PRF is robust to cases where relevant information is non-pervasive—that is, when only a subset of predictors have non-zero loadings on the relevant factors.

[4] Nothing prevents us from generalizing this to consider direct forecasts of y_{t+h} for $h \in \{1, 2, \ldots\}$—the theory is identical. For exposition's sake we deal only with y_{t+1}, knowing that $t+1$ could instead be $t+h$ but everything that follows would still hold.