# Bond Risk Premia Forecasting: A Simple Approach for Extracting Macroeconomic Information from a Panel of Indicators

Francesco Audrino* Fulvio Corsi Kameliya Filipova

University of St. Gallen

September 2012

**Abstract**

We propose a simple but effective estimation procedure to extract the level and the volatility dynamics of a latent macroeconomic factor from a panel of observable indicators. Our approach is based on a multivariate conditionally heteroskedastic exact factor model that can take into account the heteroskedasticity feature shown by most macroeconomic variables and relies on an iterated Kalman filter procedure. In simulations we show the unbiasedness of the proposed estimator and its superiority to different approaches introduced in the literature. Simulation results are confirmed in applications to real inflation data with the goal of forecasting long-term bond risk premia. Moreover, we find that the extracted level and conditional variance of the latent factor for inflation are strongly related to NBER business cycles.

**JEL classifications**: C13; C33; C53; C82; E31; E47

**Keywords**: Macroeconomic variables; Exact factor model; Kalman filter; Heteroskedasticity; Forecasting bond risk premia; Inflation measures; Business cycles

*Corresponding author. Address for correspondence: University of St. Gallen, Institute of Mathematics and Statistics, Bodanstrasse 6, CH-9000 - St. Gallen, Switzerland. Tel: +41 71 224 2431. Fax: +41 71 224 2894. Email: francesco.audrino@unisg.ch.

# 1 Introduction

In their highly influential paper, using a reduced form no–arbitrage framework with time–varying risk premia, Ang and Piazzesi (2003) conclude that macroeconomic variables have an important explanatory power for yields and that the inclusion of such variables in term structure models can improve their forecasting performances significantly. More recently, many other studies (see, among others, Ludvigson and Ng (2009b), Joslin et al. (2009), Duffee (2009) for the U.S. or Wright (2009) in an international context) have documented that macroeconomic variables capture significant predictive power for bond excess returns over and above the standard financial factors. In order to avoid relying on specific macro series, Ang and Piazzesi (2003) and Ludvigson and Ng (2009a), measure different macroeconomic fundamentals as the first principal components of blocks of large numbers of macroeconomic series.

In this paper we propose considering macroeconomic variables as possible relevant factors for modeling the dynamics of the bond risk premia process (and therefore the whole term–structure). We take into account not only the level of a macroeconomic variable, but also its volatility. Moreover, we also propose a different method for reconstructing the level and volatility dynamics of the latent macro–factor from a bunch of observable indicators. Our approach is considerably simpler from a computational perspective than the classical ones introduced in the literature and at the same time performs better in simulations as well as in a real data applications.

In macroeconomics, it is common to have a large set of indexes that measure or are highly dependent on a latent macroeconomic variable. Given the pervasiveness of heteroskedasticity in macroeconomic variables, we model the observable set of proxies using a multivariate conditionally heteroskedastic exact factor model, i.e. a linear factor model where the heteroskedastic conditional variance is a function of the past values of the latent factor (see for instance, Diebold and Nerlove 1989). In such a type of model, the conditional density, depending on unobservable variables, is generally unknown. As a consequence, the log-likelihood function cannot be obtained explicitly and hence standard maximum likelihood estimators cannot be employed (Harvey et al. 1992 ). To overcome this problem, alternative estimation procedures have been proposed in the literature: the Bayesian Markov chain Monte Carlo (MCMC) estimation methods introduced by Fiorentini et al. (2004) and the indirect inference estimators introduced by Sentana et al. (2008).

However, following the direction proposed by Diebold and Nerlove (1989) and Sentana (2004), in this study we introduce a (computationally) simple estimation approach that relies on filtering the latent factor from a panel of data via an iterated Kalman filter procedure. This approach hinges on recent results about efficient estimation of the macro-parameters in dynamic panel data models with a common factor. In particular, Gagliardini and Gourieroux (2009) showed that substituting the true factor values by their cross-sectional approximations does not lead to any asymptotic efficiency loss. For the cross–sectional reconstruction of the latent factor we propose an iterated process in which we estimate the volatility dynamics of the factor from the time series of a first (time–invariant) Kalman filter approximation of the factor and use it in a new cross–sectional conditional (time–varying) Kalman filter estimation. New volatility dynamics can be estimated from the dynamics of the new estimated factor and the procedure can be iterated until convergence.

Simulation results based on different data–generating processes and the same amount of data that are available in the empirical application show the unbiasedness of the proposed estimator for the conditional variance parameters and its superiority to other simple alternative methods, in particular, to the principal component approach used by Ludvigson and Ng (2009a).

The superiority of our approach is also confirmed by a real data application. Using a panel of 21 monthly inflation time series, we filter the level and the volatility of inflation via several different techniques. We test the ability of the estimated factors in forecasting long–term bond risk premia and find that both the level and the volatility of inflation obtained via an iterated Kalman filter significantly outperform the other competitors. Moreover, by analyzing the correspondence between the different factors and National Bureau of Economic Research (NBER) business cycles, we show that our inflation estimates are not only statistically but also economically significant.

The reminder of the paper is organized as follows. Section 2.1 describes in detail the procedure of reconstructing the level and volatility dynamics of a latent factor. Section 2.2 shows the performance of the latent macroeconomic variable and its volatility in a simulation study. In Section 3 we apply our estimation technique on real macroeconomic data. Section 4 concludes.

# 2 Reconstructing the dynamics and volatility of the latent factor

Our purpose in this section is to reconstruct the underlying time series dynamics of a latent macroeconomic variable and its volatility process from the observations of a certain number of proxies. We propose a simple estimation approach that exploits the possibility of filtering the latent factor from cross-sectional information via an iterated Kalman filter procedure.

## 2.1 Model and estimation procedure

We model the latent factor dynamics at time $t$ through a factor model for the $N$-dimensional vector of the observables $r_t = (r_{t,i})_{i=1}^N$

$$r_t = Bf_t + e_t, \qquad \text{for} \qquad t = 1...T \tag{1}$$

with $B$ the $N \times k$ matrix of factor loadings, $e_t$ the $N \times 1$ vector of idiosyncratic noises, and the latent factor $f_t$ being the variable of interest. In our empirical study of Section 3 we consider a univariate factor representing the latent inflation (i.e., using the standard notation, $f_t = \pi_t$) and, as observables, a number of index proxies for inflation such as different types of Producer and Consumer price indices.

The main assumptions of the model can be expressed in the following form:

$$\begin{pmatrix} f_t \\ e_t \end{pmatrix} |\mathbb{I}_{t-1} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Delta_t & 0 \\ 0 & \Phi \end{pmatrix}\right]. \tag{2}$$

The latent factor $f_t$ is assumed to follow a general GARCH type dynamic with $f_t|\mathbb{I}_{t-1} \sim$

$N(0, \Delta_t)$ with $\Delta_t$ a diagonal positive definite matrix of time-varying factor variances and (for identifiability) unconditional variance $\mathbb{E}[\Delta_t] = \Delta = I_k$ the identity matrix of order $k$.[1] The information set $\mathbb{I}_t$ contains current and past values of $r$ and $f$, i.e. $\mathbb{I}_t = \{r_t, f_t, r_{t-1}, f_{t-1}, \cdots,\}$. As in standard factor models, the vector of idiosyncratic noises $e_t$ is conditionally orthogonal to $f_t$ and has a positive semidefinite diagonal variance matrix $\Phi$, then the conditional distribution of $r_t$ is $r_t|\mathbb{I}_{t-1} \sim N(0, \Sigma_t)$ where $\Sigma_t = B\Delta_t B' + \Phi$ has the usual exact factor structure.

In the literature this type of model is called a multivariate conditionally heteroskedastic exact factor model and nests several models widely used in empirical finance (for instance, Diebold and Nerlove 1989). When the variance of the factor is a function of lagged values of $f_t$, as in the GARCH case, the exact form of the conditional density of $r_t$ given its past is generally unknown and, hence, the log-likelihood function cannot be explicitly obtained (Harvey et al. 1992). To overcome this problem, Bayesian Markov chain Monte Carlo (MCMC) estimation methods, simulated EM algorithm (Fiorentini et al. 2004) and indirect inference estimators (Sentana et al. 2008) have been proposed in the literature.

Here, instead, we propose a simpler approach in which we iterate between filtering the factor with a Kalman filter in the cross–sectional dimension and estimating its variance dynamics in the time series dimension. This approach hinges on the idea contained in the recent literature on estimators of the macro-parameters in dynamic panel data models with

[1] Although possible in principle to extend the model to include dynamics in the conditional mean of the factor, this would certainly complicate both the reconstruction of the latent factor and the estimation of the dynamics of $\Delta_t$. Since our purpose in this paper is to propose an unbiased estimation method which is as simple as possible, we leave this extension of the model for future research.

a common factor where the macro-parameter is estimated by means of cross-sectional approximations (Forni and Reichlin 1998,Forni et al. 2004, Gagliardini and Gourieroux 2009 ). These studies show that, under certain speed of convergence assumptions,[2] estimating the macro-parameter on the cross-sectional approximations of the factors is root–T consistent, asymptotically normal and achieves the same asymptotic efficiency bound as the one obtained with an observable factor (i.e. the Cramer-Rao bound in linear Gaussian models). Therefore, the estimators built on the approximated factor are asymptotically equivalent to the unfeasible estimator that uses the true factor values. These efficiency results are obtained under certain asymptotic schemes which are not expected to necessary hold in our setting. Therefore, whenever these asymptotic conditions are not satisfied, estimators based on more complex simulated estimation methods (as the ones in Fiorentini et al. 2004 and Sentana et al. 2008) are expected to be asymptotically more efficient. However, the big advantage of the proposed estimator is to be computationally much simpler. This advantage is due to the way the proposed estimator effectively exploit the cross–sectional dimension (to reconstruct the factor) in combination with the time–series information (used to filter the variance of the factor).

Different approaches can be used to approximate $f_t$: simple cross sectional averaging, principal component analysis (PCA) or factor analysis (FA). In this study we propose a reconstruction of the $f_t$ factor by an iterative procedure in which the factor is first estimated with a Kalman filter using the cross-section of the observable indicators at our disposal. From the time series of this first approximation of the factor, the variance dynamics are

[2]When $N, T \to \infty$ and $T/N \to c > 0$ the fixed effects estimator is consistent, while if $N, T \to \infty$ such that $T^b/N = O(1), b > 1$ the estimator is efficient.

estimated in a classical GARCH framework. The estimated GARCH dynamics of the factor conditional variance are then used in a conditional Kalman filter estimation to obtain new factor estimates. This iterative procedure is run until convergence. Although we apply this approach to a case where a one factor model arises naturally, this procedure could be directly extended to the case of multiple factors provided that one is not interested in the exact identification of the different factors (because of the indeterminacy induced by factor rotation).

Before starting the procedure, we need an estimate of the factor loading matrix $B$. Given that in these types of models the factor loadings are assumed to be constant over time, they can be conveniently estimated from unconditional quantities. Moreover, conditionally heteroskedastic factor models also imply unconditional covariance matrices that have an exact $k$ factor structure as in the traditional factor models. Hence, recalling that $\Delta = I_k$, the unconditional covariance matrix $\Sigma$ can be written as

$$\Sigma = BB' + \Phi. \tag{3}$$

Given the different scale of the indices (which have different units of measures), it is desirable to standardize the variable to avoid the problem of having one variable with a large variance unduly influencing the determination of the factor loadings. Standardizing by the individual volatility and working with the correlation matrix is then a customary choice. Clearly, the correlation matrix $R = D^{-1}\Sigma D'^{-1}$ with $D = \text{diag}(\Sigma)$ will also have the same factor structure

$$R = B^* B^{*'} + \Phi^* \tag{4}$$

with $B^* = D^{-1}B$ and $\Phi^* = D^{-1}\Phi D'^{-1}$.

Since in our case all the observed indexes are mainly driven by a single latent macroeconomic variable they are supposed to measure, we assume a factor structure with only one common factor (i.e. $k = 1$). Then, the correlation matrix takes the following simple structure.

$$R = \begin{bmatrix} 1 & b_1^* b_2^* & \dots & b_1^* b_N^* \\ b_2^* b_1^* & 1 & \dots & b_2^* b_N^* \\ \vdots & & \ddots & \vdots \\ b_N^* b_1^* & b_N^* b_2^* & \dots & 1 \end{bmatrix}$$

where $[B^*]_i = b_i^*$ is the generic element of the $N \times 1$ vector $B^*$. This structure, together with the fact that the factor loadings of the proxy are assumed to be all positive, suggests the possibility to estimate the vector of standardized factor loadings $B^*$ by simply minimizing the difference between any generic off diagonal element of the matrix $B^* B^{*'}$ with the corresponding element of the sample unconditional correlation matrix $[S^*]_{ij} = s_{ij}^*$, that is

$$\hat{b}^* = \underset{b^*}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{j \neq i} (b_i^* b_j^* - s_{i,j}^*)^2. \qquad s.t. \quad 0 < b_i^* < 1 \qquad \forall i \tag{5}$$

The minimization algorithm in (5) projects the sample correlation matrix into the space spanned by single factor models.

Having the estimated standardized factor loadings $\hat{B}^*$'s, we can estimate the elements of the diagonal matrix $\Phi^*$ as $[\hat{\Phi}^*]_{ii} = 1 - (\hat{b}_i^*)^2$. Then the original idiosyncratic variance matrix and factor loadings are simply obtained as $\hat{\Phi} = \hat{D} \hat{\Phi}^* \hat{D}'$ and $\hat{B} = \hat{D} \hat{B}^*$ respectively.

With $\hat{B}$ and $\hat{\Phi}$ at hand, we can now start the Kalman filter iteration. If the joint conditional distribution of $r_t$ and $f_t$ given $\mathbb{I}_{t-1}$ is normal, the model (1) has a natural time–series state–space representation. In fact, considering the common factor $f_t$ as state variable, equation (1) could be seen as a standard measurement equation. When $\Delta_t$ is considered as a given observable, the Kalman filter would coincide with the conditional expectation of $f_t$ given $r_t$ and $\Delta_t$, i.e. $E[f_t|r_t, \Delta_t]$, which is optimal in the conditional mean squared error sense.[3] Thus, the conditional Kalman filter estimate of the common factor would be given by the (unfeasible) updating equation of the filter

$$f_t^{CK} = \Delta_t B' \Sigma_t^{-1} r_t = \Delta_t B'(B \Delta_t B' + \Phi)^{-1} r_t. \tag{6}$$

This estimator can be seen as a Bayesian approach for the cross–sectional estimation of the factor. More precisely, the unfeasible estimator in (6) corresponds to the mean of the posterior distribution of $f_t$ given the data $r_t$ in a Bayesian approach that considers $f_t$ as a random variable with prior distribution $f_t \sim N(0, \Delta_t)$.[4]

In order to have a feasible conditional Kalman filter, we propose to start the iterative procedure from the following filter with time–invariant weights

$$\hat{f}_t^{(0)} = \hat{B}' \hat{\Sigma}^{-1} r_t = \hat{B}'(\hat{B}\hat{B}' + \hat{\Phi})^{-1} r_t \tag{7}$$

using the estimates $\hat{B}$ and $\hat{\Phi}$ obtained from the unconditional information.

Having this first reconstruction of the dynamics of the latent macro–variable, we then get an estimate of the dynamics of its volatility by estimating a GARCH model on $\hat{f}_t$. In

[3] Actually, the optimality of the Kalman filter extraction of the factor holds under the more general assumption that $f_t$ and $r_t$ follow a conditional joint distribution that is elliptically symmetric (Sentana 1991).

[4] Hence, our simplifying assumption of a zero mean prior implies that the information in the past dynamics of the conditional mean is not considered in the forecast of the conditional mean of the factor, but, obviously, does not prevent the posterior mean to be different from zero (as typically shown in empirical data).

this way we obtain a first estimate of the dynamics of the conditional variance of the factor i.e. $\hat{\Delta}_t^{(0)}$ which is then used in the conditional Kalman filter estimation of the factor

$$\hat{f}_t^{(1)} = \hat{\Delta}_t^{(0)} \hat{B}' \hat{\Sigma}_t^{-1} r_t = \hat{\Delta}_t^{(0)} \hat{B}' \left( \hat{B} \hat{\Delta}_t^{(0)} \hat{B}' + \hat{\Phi} \right)^{-1} r_t \tag{8}$$

from which a new reconstruction of the latent factor can be computed and a new conditional variance dynamics $\hat{\Delta}_t^{(1)}$ estimated. Iterating this procedure provides our proposed estimator for the dynamics of the latent factor and its conditional variance. Note that in practice, only a small number of iterations is necessary to reach converge and the algorithm is very fast.

## 2.2 Simulations

We first judge the performance of the proposed approach on the accuracy in the reconstruction of the time series of the latent factor $f_t$. The first employed data generating process (DGP) is a one factor model with the latent factor following a GARCH type dynamics with zero mean and unconditional unit variance. We simulate 1000 paths and for each path we assume 49 years of monthly observations ($T = 588$). Similarly to our real data application, we assume to have 20 observable indicators for the latent macroeconomic variable ($N = 20$). The true $\beta$s in the DGP are randomly chosen within a range of values analogous to that estimated on the empirical data (see data Appendix).

For comparison purposes we also include the result obtained with a simple cross–sectional average of the indexes, the factor score obtained with cross-sectional OLS regression, the PCA and the FA with one factor. When $N$ is large enough (so that idiosyncratic errors are diversified away) we have that the simple cross-sectional average is

$\bar{r}_t \simeq \left(\frac{1}{N}\sum_{i=1}^{N} b_i\right) f_t$; thus the factor values are recovered up to a scaling constant. We account for this scaling constant by simply dividing the series of the cross-sectional averages $\bar{r}_t$ by $\frac{1}{N}\sum_{i=1}^{N} \hat{b}_i$. The cross–sectional OLS regression is another common method to generate factor scores. Contrary to our Kalman filter approach which consider $f_t$ as a random variable, this approach assume the factor to be an unknown parameter and estimate it by the cross–sectional regression $f_t^{OLS} = (\hat{B}'\hat{B})^{-1}\hat{B}'r_t$. The FA is performed using the Matlab command "factoran" which performs maximum likelihood estimate of the factor loadings and computes factor scores using the weighted least–square (or Barlett) method (which also treats the factor scores as fix parameters).

Given that the OLS regression approach completely discards the information contained in $\Delta_t$ and $\Phi$, while FA neglects the information contained in the dynamics of $\Delta_t$, we expect them to be less efficient than the Iterated Kalman filter method who optimally exploits the information in both $\Phi$ and $\Delta_t$.

To judge the accuracy in reconstructing the $f_t$ series with the various approaches, we compute the Root Mean Square Error (RMSE) for each simulated path between the true path of the latent factor and the estimated one. For each simulation path we also compute the correlation coefficient between the two series. Results are reported in the first two rows of Table 6 Panel A.

[Table 1 about here.]

According to both metrics, our proposed procedure for the latent $f_t$ process turns out to be the most precise; it is the one with, on average, the smallest RMSE and the highest correlation coefficient.

We then evaluate the ability of the different approaches to reconstruct the volatility dynamics of the true factor by computing the RMSE and correlation coefficient between the true series of simulated volatilities and the reconstructed ones obtained by fitting a GARCH(1,1) process to the estimated $f_t$ series. Again, the Iterated Kalman filter provides the reconstruction of the latent factor volatility with, on average, the lowest RMSE and the highest correlation coefficient, as shown in the last two rows of Table 6 Panel A.

We notice that the improvements in the RMSE are typically more pronounced than those on correlations. This can be explained by the fact that correlations only considers common directions in the movements of two variables while completely neglecting the effective “distance” between the two while the RMSE is very sensitive to increases in the estimation errors (over- or under-estimations). Given that both PCA and FA ignore dynamics in $\Delta_t$ they can in some periods overestimate and in other underestimate the variable of interest thus increasing the overall estimation error and hence the RMSE compared to the Kalman filter approach which, on the contrary, optimally utilizes the information contained in both $\Phi$ and $\Delta_t$.

Finally, in Figure 1, Panel A, we plot the distributions of the estimated parameters of the GARCH process for the volatility.

[Figure 1 about here.]

The figure clearly shows that the estimates of the true parameters $\alpha$ and $\beta$ of the GARCH process in the factor DGP are both unbiased and reasonably accurate. Similar results are obtained for simulations with slightly different values of the GARCH parameters (available from the authors upon request).

## 2.3 Robustness checks

In this section we provide several robustness checks of our procedure by testing it on two more challenging volatility processes (with a purposely misspecified DGP) and with fewer time-series data points. In the first DGP the diagonal variance matrix of the idiosyncratic noise $\Phi$, which was kept constant over time in the previous standard set up, is now also time–varying, with each idiosyncratic component following a different GARCH process. The objective of this simulation exercise is to test the robustness of our procedure in a misspecified set up featuring GARCH dynamics in both the factor and idiosyncratic conditional variances, i.e. with time–varying $\Delta_t$ and $\Phi_t$.

The second DGP consists of a two-regime process with lagged return as the threshold variable where the local conditional variance evolves according to a FIGARCH(1,d,1) model (see Baillie et al. 1996) in one regime and a model that is not of a GARCH type in the second regime. Results for the two more complex volatility DGPs are reported in Table 7.

[Table 2 about here.]

The results are qualitatively similar, although quantitatively worse, to those previously obtained in the correctly specified set up, confirming in both cases the more accurate reconstruction of the latent process by the proposed iterated Kalman filter method. Differences with respect to the accuracy of the reconstructed factor volatility series are particularly evident: This result shows the main peculiarity of the proposed approach, that is its ability to deal with the autoregressive conditional heteroscedasticity present in the data. The alternative methods neglecting this aspect, may behave reasonably well for the reconstruction of the factor level, but completely fail in capturing the correct factor volatility. As expected, differences across the models become less evident, in particular in the second DGP, given that all methods used are significantly misspecified.

As for the previous standard case, we plot in Figure 1, Panel B the distribution of the $\alpha$ and $\beta$ parameter estimates in the case of DGP process with time varying (GARCH type) idiosyncratic noise $\Phi_t$. GARCH parameter estimates seem to remain unbiased even in this misspecified context.

[Figure 1 about here.]

Finally, we test the sensitiveness of our procedure to shorter datasets by performing simulations with data samples spanning 15 and 30 years (instead of the 49 years employed previously). For space concerns, Table 8 reports the results of these stress tests only for the most challenging DGP i.e. the one having a GARCH dynamics with two regimes (the similar results obtained with the others DGP are available from the authors upon request).

[Table 3 about here.]

Although some deterioration is observed in the estimation performances of the conditional variances (as should be expected), the general performance in the reconstruction of the factor and the relative ranking remain substantially unchanged.

# 3 Real data application: bond risk premia forecasting

Economic theory suggests that (a great portion of) bond term premia variation is driven by macroeconomic fundamentals. Yet, the link between macroeconomic activity and risk premia might be hard to detect. Using different modeling setups, many recent studies (see, among others, Ludvigson and Ng (2009b), Joslin et al. (2009), or Duffee (2009)) document that macroeconomic variables capture significant predictive power for excess returns over and above the standard financial factors. In this section we assess the performance of our iterated Kalman filter technique in forecasting long–term bond excess returns in comparison with principal components analysis and factor analysis.[5] Results obtained applying a principal components analysis are not reported in Section 3.3 given that they are qualitatively the same, but slightly worse, than those obtained using factor analysis.

## 3.1 Data and estimated inflation levels and variances

In our empirical study two different datasets are used.

*Bond Data*

We use monthly data (June 1961 onward) from the Federal Reserve Board constructed as in Gürkaynak et al. (2006).[6] Bond excess returns are calculated in the classical way as 1-year holding period returns in excess of the one–year risk–free rate.

Furthermore, we construct our tent–shape bond–return forecasting factor described in Cochrane and Piazzesi (2005) (hereafter CP factor) as a linear combination of forward rates. The inclusion of the CP factor is motivated simply by the fact that it has high explanatory power for bond excess returns beyond the one that is captured by the yield curve "level," "slope," and "curvature" factors.

*Macroeconomic Data*

[5]Another possible alternative procedure is the one proposed by Harvey et al. (1992). Given the dimensionality of the problem, however, that approach is too computational expensive and is not implemented.

[6]The data are available under http://www.federalreserve.gov/econresdata/researchdata.htm.

The second dataset consists of monthly observations for 21 U.S. inflation time series. Exact description of the data is given in Appendix A. The panel spans the period January 1959 – December 2007 and has already been used as a part of other studies: see, among others, Stock and Watson (2005), Ludvigson and Ng (2009b) and Ludvigson and Ng (2009a). We build two alternative pairs of estimates for inflation levels and variances. First, similar to Ludvigson and Ng (2009a), we extract both the first principal component (PCA) and the first factor (FA) as measures for inflation's level. PCA and FA volatility are computed from fitting a GARCH(1,1) to the estimated principal component and the estimated factor, respectively. Our second approach for reconstructing the level and the variance of inflation is based on the iterated Kalman filter procedure described in Section 2.1.

For our analysis we take the largest common period of the two datasets and split it into two parts. We consider June 1961 to December 2003 as in-sample period. The rest of the data (January 2004 - December 2008) has been left to evaluate the out-of-sample forecasting performance of the different predictors. Summary statistics of the data are reported in Table 1.

The adequacy of the one factor structure may be questionable. In fact, the assumption of the one factor structure is primarily given by the economic consideration that all the variable in the data set are all proxy of the same underlying macroeconomic variable i.e. inflation. From a statistical point of view we can observe that the first principal component explain about 53% of the total variance of the dataset while all the other components are below 10%. The presence of a highly persistent heteroscedasticity in the series, which justifies the use of a GARCH(1,1) model, is given by the highly significant results of the Engle ARCH test with a large number of lags of 50.

| | $rx^{(5)}$ | $rx^{(10)}$ | $rx^{(20)}$ | $rx^{(30)}$ | CP | $\pi^{IK}$ | $\mathtt{vol}\pi^{IK}$ | $\pi^{PCA}$ | $\mathtt{vol}\pi^{PCA}$ | $\pi^{FA}$ | $\mathtt{vol}\pi^{FA}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Panel A: | | | | | | | | | | | |
| Mean | 0.011 | 0.013 | 0.011 | 0.009 | 0.006 | 0.838 | 1.034 | 0.034 | 0.876 | 0.497 | 0.499 |
| Std | 0.056 | 0.104 | 0.198 | 0.325 | 0.019 | 0.579 | 2.109 | 1.014 | 0.694 | 0.401 | 0.370 |
| AC1 | 0.931 | 0.921 | 0.880 | 0.798 | 0.916 | 0.989 | 0.993 | 0.973 | 0.953 | 0.991 | 0.990 |
| Panel B: | | | | | | | | | | | |
| $rx^{(5)}$ | 1.00 | | | | | | | | | | |
| $rx^{(10)}$ | 0.96 | 1.00 | | | | | | | | | |
| $rx^{(20)}$ | 0.82 | 0.92 | 1.00 | | | | | | | | |
| $rx^{(30)}$ | 0.62 | 0.72 | 0.90 | 1.00 | | | | | | | |
| CP | 0.43 | 0.48 | 0.49 | 0.44 | 1.00 | | | | | | |
| $\pi^{IK}$ | -0.22 | -0.25 | -0.26 | -0.26 | -0.06 | 1.00 | | | | | |
| $\mathtt{vol}\pi^{IK}$ | -0.31 | -0.36 | -0.31 | -0.29 | -0.15 | 0.89 | 1.00 | | | | |
| $\pi^{PCA}$ | -0.30 | -0.29 | -0.30 | -0.29 | -0.41 | 0.55 | 0.42 | 1.00 | | | |
| $\mathtt{vol}\pi^{PCA}$ | -0.23 | -0.24 | -0.26 | -0.26 | -0.38 | 0.49 | 0.49 | 0.69 | 1.00 | | |
| $\pi^{FA}$ | -0.27 | -0.29 | -0.30 | -0.30 | -0.13 | 0.97 | 0.85 | 0.68 | 0.61 | 1.00 | |
| $\mathtt{vol}\pi^{FA}$ | -0.26 | -0.30 | -0.30 | -0.30 | -0.11 | 0.97 | 0.89 | 0.53 | 0.55 | 0.98 | 1.00 |
| NBER | | | | | 0.04 | 0.46 | 0.45 | 0.17 | 0.24 | 0.43 | 0.41 |

Table 1: Panel A reports summary statistics for the following variables: 5, 10, 20, 30 year bond excess returns (denoted by $rx^{(5)}$, $rx^{(10)}$, $rx^{(20)}$, $rx^{(30)}$, respectively), Cochrane and Piazzesi (2005) factor (denoted by CP), inflation level and inflation volatility factors estimated by iterated Kalman filter (denoted by $\pi_t^{IK}$ and $\mathtt{vol}\pi_t^{IK}$),by factor analysis (denoted by $\pi_t^{FA}$ and $\mathtt{vol}\pi_t^{FA}$), and by principal components analysis (denoted by $\pi_t^{PCA}$ and $\mathtt{vol}\pi_t^{PCA}$). NBER is a binary variable, where one indicates month designated as recessions by the National Bureau of Economic Research. AC1 denotes the first autocorrelation coefficient. Panel B reports cross–correlations.

Figure 2 illustrates the difference between the inflation's level and the inflation's volatility factors obtained using the three different techniques.

[Figure 2 about here.]

As Figure 2 clearly shows, the estimated inflation's level and volatility factors obtained from the three competing approaches are significantly different.

## 3.2 Financial variables, inflation measures, and business cycles

To begin with, we analyze the correspondence between the NBER business cycles and the different financial and inflation measures. The last row of Table 1 reports the results. The weak correlation (around 0.04) between the NBER recession and CP factor confirms Ludvigson and Ng (2009b) finding that, without macro factors, bond risk premia appear virtually acyclical. Yet, theory says that risk premia have a marked counter–cyclical behavior, compensating the investors for macroeconomic risks. The almost two times higher correlation between the NBER business cycles indicator and the iterated Kalman filter inflation variables as well as the ones obtained from FA in comparison to those estimated with the PCA approach assures more pronounced cyclical fluctuations in bond risk premia. By its iterated nature, our measures for inflation seem to better capture perceptions of risks looming on the investors horizon. Thus, they convey valid and timely information over and above that contained in other financial and PCA inflation fundamentals. These findings make the inflation factors obtained by the iterated Kalman filter approach highly economically significant. No particularly significant difference in the correlations with NBER business cycles can be seen between our procedure and a classical FA.

## 3.3 Long–term bond risk premia forecasting results

### 3.3.1 Predictive regressions

To assess the in-sample performance of our procedure in comparison with the alternative approaches, on a first step, we investigate the impact of the different pairs of inflation factors (i.e. level and volatility) as predictors for bond excess returns at different maturities. We find that among all inflation variables our inflation volatility factor explains the highest portion of the risk premia across all maturities. Interestingly, the model with the two inflation level factors yields the highest adjusted $R^2$. Hence, our iterated Kalman technique is able to uncover important piece of inflation level information, not captured by the conventional PCA and FA techniques. Details can be found in Appendix B.

Given the fact that the impact of inflation on bond risk premia is comparatively small, following the term structure literature, we repeat the analysis above including the Cochrane and Piazzesi factor in the regressions. The reason for that is that the Cochrane and Piazzesi factor summarizes all the information in individual yield spreads and forward spreads that had been proven to contain high predictive power for bond excess returns (see for example Ludvigson and Ng (2009b), Cochrane and Piazzesi (2005)). To this goal we run the following regressions:

$$\text{Model } \mathcal{M}_1: \quad rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 CP_t + \varepsilon_{t+12}^{(n)}$$

$$\text{Model } \mathcal{M}_2: \quad rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 CP_t + \gamma_2 \pi_t^{IK} + \varepsilon_{t+12}^{(n)}$$

$$\text{Model } \mathcal{M}_3: \quad rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 CP_t + \gamma_2 \mathtt{vol}\pi_t^{IK} + \varepsilon_{t+12}^{(n)}$$

$$\text{Model } \mathcal{M}_4: \quad rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 CP_t + \gamma_2 \pi_t^{FA} + \varepsilon_{t+12}^{(n)}$$

$$\text{Model } \mathcal{M}_5: \quad rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 CP_t + \gamma_2 \mathtt{vol}\pi_t^{FA} + \varepsilon_{t+12}^{(n)}$$

$$\text{Model } \mathcal{M}_6: \quad rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 CP_t + \gamma_2 \pi_t^{IK} + \gamma_3 \mathtt{vol}\pi_t^{IK} + \varepsilon_{t+12}^{(n)}$$

$$\text{Model } \mathcal{M}_7: \quad rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 CP_t + \gamma_2 \pi_t^{FA} + \gamma_3 \mathtt{vol}\pi_t^{FA} + \varepsilon_{t+12}^{(n)},$$

$$\text{Model } \mathcal{M}_8: \quad rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 CP_t + \gamma_2 \pi_t^{IK} + \gamma_3 \pi_t^{FA} + \varepsilon_{t+12}^{(n)}$$

$$\text{Model } \mathcal{M}_9: \quad rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 CP_t + \gamma_2 \mathtt{vol}\pi_t^{IK} + \gamma_3 \mathtt{vol}\pi_t^{FA} + \varepsilon_{t+12}^{(n)},$$

where $rx_{t+12}^{(n)}$ are the excess returns on an $n$ year nominal bond ($n = 5, 10, 20, 30$) at time $t + 12$. $CP_t$ represents the CP factor, $\pi_t$ and $\mathtt{vol}\pi_t$ denote the inflation level and inflation volatility factors, estimated by the two different approaches: iterated Kalman filter (denoted by $\pi_t^{IK}$ and $\mathtt{vol}\pi_t^{IK}$) and factor analysis (denoted by $\pi_t^{FA}$ and $\mathtt{vol}\pi_t^{FA}$), respectively. To this end, we estimate nine different models. First, we regress the excess returns only on CP factor (Model $\mathcal{M}_1$). This regression should serve as a benchmark model. Then, in Model $\mathcal{M}_2$ and Model $\mathcal{M}_3$ we add one more predictor, the level and the volatility of inflation, each estimated by the iterative Kalman filter approach. We repeat the same procedure for the next two models (Model $\mathcal{M}_4$ and Model $\mathcal{M}_5$), where we add once again the level and the volatility of inflation, this time estimated by the FA technique. In Model $\mathcal{M}_6$ and Model $\mathcal{M}_7$ we take into consideration all three predictors: CP factor, level and volatility of inflation. The only difference between Model $\mathcal{M}_6$ and Model $\mathcal{M}_7$ is in the way the inflation variables are measured. In particular, in Model $\mathcal{M}_6$ the inflation variables are derived by the iterated Kalman filter procedure, whereas in $\mathcal{M}_7$ FA has been used. In contrast to the previous models, where the main idea is to assess performance, the individual filtering techniques, the last two models (Model $\mathcal{M}_8$ and Model $\mathcal{M}_9$) provide a direct comparison between the two level (Model $\mathcal{M}_8$) and the two volatility (Model $\mathcal{M}_9$) factors. All coefficients are estimated with ordinary least squares, and standard errors are corrected for autocorrelation and heteroskedasticity.[7] Table 2 and Table 3 present the results.

The estimated coefficients for the CP factor are positive and highly significant for predicting bond risk premia at all maturities. Fully in line with the literature, the CP factor accounts for around 28% of the excess returns variation. The strength of the predictive power of the inflation factors changes with time to maturity of a bond, explaining up to 6% of the variation in addition to the CP factor. The estimated coefficients for level and volatility of inflation are negative, and they are significant most of the time. The negative correlation between the different inflation measures and excess returns is quite intuitive, as higher inflation decreases the value of the nominal bond. Including both level and volatility of the inflation factor (see Models $\mathcal{M}_6$ and $\mathcal{M}_7$) in the regression does not seem to improve the accuracy, and both predictors become statistically not significant.[8]

[7]In particular, we follow Ludvigson and Ng (2009b) and Cochrane and Piazzesi (2005) and compute $t$-statics using the Newey-West adjustment with 18 lags.

[8]This result is a consequence of the high correlation between the two variables (and both series are very persistent) together with the necessary Newey-West correction that substantially lowers the t-statistics.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Intercept | $-0.002$ | 0.013 | 0.005 | 0.013 | 0.013 | 0.004 | 0.011 | 0.003 | 0.008 |
| | $(-0.310)$ | (1.168) | (0.678) | $(-0.173)$ | (0.185) | (0.339) | (1.118) | ( 0.724) | (0.824) |
| CP Factor | $1.533^{***}$ | $1.474^{***}$ | $1.410^{***}$ | $1.430^{***}$ | $1.444^{***}$ | $1.408^{***}$ | $1.420^{***}$ | $1.3200^{***}$ | $1.418^{***}$ |
| | (4.749) | (4.555) | (4.219) | (3.982) | (4.333) | (3.992) | (4.294) | (4.016) | (4.080) |
| Inflation Level | | $-0.027^{*}$ | | | | -0.060 | | 0.068 | |
| (Iterated Kalman) | | $(-1.756)$ | | | | (-1.131) | | (1.599) | |
| Inflation Vol | | | $-0.028^{*}$ | | | $-0.005$ | | | $-0.003$ |
| (Iterated Kalman) | | | $(-1.650)$ | | | $(-0.826)$ | | | $(-0.474)$ |
| Inflation Level | | | | $-0.009$ | | | $-0.060$ | $-0.124^{*}$ | |
| (FA) | | | | $(-1.430)$ | | | $(-1.312)$ | $(-2.028)$ | |
| Inflation Vol | | | | | 0.035 | | 0.006 | | -0.012 |
| (FA) | | | | | (0.602) | | ( 0.615) | | (-0.398) |
| $R^2$ | 0.257 | 0.285 | 0.294 | 0.287 | 0.293 | 0.294 | 0.296 | 0.316 | 0.312 |
| Adjusted $R^2$ | 0.256 | 0.283 | 0.292 | 0.285 | 0.290 | 0.290 | 0.291 | 0.296 | 0.292 |

PANEL B: PREDICTIVE REGRESSION ANALYSIS 10 YEAR EXCESS RETURNS

| | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_6$ | $\mathcal{M}_7$ | $\mathcal{M}_8$ | $\mathcal{M}_9$ |
|---|---|---|---|---|---|---|---|---|---|
| Intercept | $-0.013$ | 0.018 | 0.001 | 0.017 | 0.019 | 0.002 | 0.017 | 0.005 | 0.011 |
| | $(-1.164)$ | ( 0.991) | (0.088) | (1.052) | (1.097) | (0.121) | (0.983) | (0.781) | (0.556) |
| CP Factor | $3.025^{***}$ | $2.899^{***}$ | $2.772^{***}$ | $2.817^{***}$ | $2.840^{***}$ | $2.778^{***}$ | $2.817^{***}$ | $2.841^{***}$ | $2.792^{***}$ |
| | (4.649) | (4.521) | (4.251) | (4.379) | (4.457) | (4.203) | (4.336) | (4.177) | (4.243) |
| Inflation Level | | $-0.035^{*}$ | | | | $-0.002$ | | 0.087 | |
| (Iterated Kalman) | | $(-1.751)$ | | | | $(-0.078)$ | | (1.186) | |
| Inflation Vol | | | $-0.011^{*}$ | | | $-0.010$ | | | $-0.006$ |
| (Iterated Kalman) | | | $(-1.917)$ | | | $(-0.960)$ | | | $(-0.562)$ |
| Inflation Level | | | | $-0.056^{*}$ | | | $-0.057$ | $-0.179^{*}$ | |
| (FA) | | | | $(-1.990)$ | | | $(-.581)$ | $(-1.709)$ | |
| Inflation Vol | | | | | $-0.059^{*}$ | | 0.001 | | $-0.029$ |
| (FA) | | | | | $(-1.941)$ | | ( 0.011) | | $(-0.572)$ |
| $R^2$ | 0.283 | 0.320 | 0.328 | 0.327 | 0.327 | 0.328 | 0.328 | 0.338 | 0.330 |
| Adjusted $R^2$ | 0.282 | 0.317 | 0.325 | 0.324 | 0.325 | 0.324 | 0.324 | 0.334 | 0.326 |

Table 2: Results for ordinary least squares regressions for nine different models (labeled as $\mathcal{M}_1$, $\mathcal{M}_2$,. . .,$\mathcal{M}_9$) utilizing annual returns on 5- and 10-year Treasury bonds. Standard errors are corrected for autocorrelation and heteroskedasticity. t-statistics are reported in parenthesis. Asterisks *

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Intercept | $-0.042^{**}$ | 0.017 | $-0.016$ | 0.014 | 0.020 | $-0.005$ | 0.019 | -0.005 | 0.013 |
| | $(-1.976)$ | (0.634) | $(-0.972)$ | (0.639) | (0.784) | $(-0.157)$ | (0.730) | (-0.159) | (0.391) |
| CP Factor | $5.901^{***}$ | $5.659^{***}$ | $5.438^{***}$ | $5.507^{***}$ | $5.540^{***}$ | $5.482^{***}$ | $5.533^{***}$ | $5.270^{***}$ | $5.498^{***}$ |
| | (4.366) | (4.439) | (4.300) | (4.368) | (4.425) | (4.298) | (4.354) | (4.216) | (4.287) |
| Inflation Level | | $-0.066^{*}$ | | | | $-0.019$ | | 0.147 | |
| (Iterated Kalman) | | $(-1.964)$ | | | | $(-0.362)$ | | (1.186) | |
| Inflation Vol | | | $-0.019^{**}$ | | | $-0.015$ | | | $-0.005^{**}$ |
| (Iterated Kalman) | | | $(-2.348)$ | | | $(-1.169)$ | | | $(-0.378)$ |
| Inflation Level | | | | $-0.106^{*}$ | | | $-0.018$ | $-0.314^{*}$ | |
| (FA) | | | | $(-2.268)$ | | | $(-0.139)$ | $(-1.841)$ | |
| Inflation Vol | | | | | $-0.116^{*}$ | | $-0.097$ | | $-0.090$ |
| (FA) | | | | | $(-2.248)$ | | $(-0.637)$ | | $(-1.043)$ |
| $R^2$ | 0.299 | 0.336 | 0.340 | 0.341 | 0.342 | 0.342 | 0.343 | 0.351 | 0.346 |
| Adjusted $R^2$ | 0.297 | 0.334 | 0.338 | 0.339 | 0.339 | 0.337 | 0.339 | 0.348 | 0.342 |

Panel B: Predictive Regression Analysis: 30 Year Excess Returns

| | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_6$ | $\mathcal{M}_7$ | $\mathcal{M}_8$ | $\mathcal{M}_9$ |
|---|---|---|---|---|---|---|---|---|---|
| a Intercept | $-0.072^{**}$ | 0.029 | $-0.031$ | 0.024 | 0.034 | 0.008 | 0.035 | $-0.007$ | $-0.040$ |
| | $(-2.027)$ | (0.798) | $(-1.215)$ | (0.730) | (0.792) | (1.016) | (0.946) | $(-0.193)$ | $(-0.368)$ |
| CP Factor | $8.779^{***}$ | $8.372^{***}$ | $8.057^{***}$ | $8.118^{***}$ | $7.874^{***}$ | $8.213^{***}$ | $8.179^{***}$ | $7.727^{***}$ | $8.203^{***}$ |
| | (3.830) | (4.013) | (3.911) | (3.974) | (4.023) | (3.886) | (3.966) | (3.815) | (3.928) |
| Inflation Level | | $-0.112^{**}$ | | | | $-0.070$ | | 0.243 | |
| (Iterated Kalman) | | $(-2.292)$ | | | | $(-0.808)$ | | (1.370) | |
| Inflation Vol | | | $-0.030^{***}$ | | | $-0.013$ | | | 0.004 |
| (Iterated Kalman) | | | $(-2.602)$ | | | $(-0.751)$ | | | (0.233) |
| Inflation Level | | | | $-0.178^{***}$ | | | 0.030 | $-0.522$ | |
| (FA) | | | | $(-2.604)$ | | | (0.169) | $(-2.059)$ | |
| Inflation Vol | | | | | $-0.197^{***}$ | | $-0.023$ | | $-0.220$ |
| (FA) | | | | | $(-2.628)$ | | $(-1.145)$ | | $(-1.577)$ |
| $R^2$ | 0.246 | 0.285 | 0.283 | 0.289 | 0.287 | 0.286 | 0.290 | 0.298 | 0.296 |
| Adjusted $R^2$ | 0.244 | 0.282 | 0.280 | 0.287 | 0.284 | 0.282 | 0.287 | 0.295 | 0.291 |

Table 3: Results for ordinary least squares regressions for nine different models (labeled as $\mathcal{M}_1$, $\mathcal{M}_2$,…,$\mathcal{M}_9$) utilizing annual returns on 20- and 30-year Treasury bonds. Standard errors are corrected for autocorrelation and heteroskedasticity. t-statistics are reported in parenthesis. Asterisks $^*$

### 3.3.2 Out-of-sample predictability

Although, at first glance, both filtering techniques seem to perform equally well in this in-sample investigation, the ability of our approach to reconstruct in a more accurate way both the level and the volatility of inflation has empirical merits out-of-sample. To support this, we run a genuine out-of-sample experiment for the remaining period in our sample. Forecasting results covering the period January 2004 to December 2008 are shown in Table 9.

[Table 4 about here.]

The superior predictive ability tests of Hansen (2005) (see Table 9) reveal that our inflation's level and volatility measures on top of the CP factor matter for forecasting bond risk premia, significantly outperforming other alternatives. Similar conclusions can be drawn by using the Model Confidence Set (hereafter MSC) tests of Hansen et al. (2011). MCS results strongly support the accuracy and the better performance of our approach over the competitors for the period 2004-2008. More precisely, in terms of mean squared error, Model $\mathcal{M}_3$ and Model $\mathcal{M}_6$ dominate at 10%, 5 % and 5 % for the 10-, 20- and 30–year bond risk premia, respectively. The results persist in the context of mean absolute error: Model $\mathcal{M}_3$ and Model $\mathcal{M}_6$ are superior at 10% and 5 % for the 10- and 20–year bond excess returns, respectively. No particular model dominates for the 5–year bond risk premia.

The stability and strength of these results is noteworthy given the out–of–sample sample period we consider. Using data from 2004:01 to 2008:12, we capture interesting times during which inflation tends to reverse sign and become negative. At that time the economy was headed for a recession and the long-term interest rates refused to follow short-term interest rates up as the Fed tightened monetary policy. In the term structure of interest rates literature (part of) this period is known as "conundrum" and has turned out difficult to forecast. Thus, our iterated Kalman factors comprises important information about the inflation looming over investor's horizon that cannot be extracted by the FA and PCA.

As a robustness check, we repeat the analysis over different sub samples. The series of conducted in-sample regressions show that the estimated coefficients remain stable and significant across bond maturities.

Given the fact that inflation has different impact on bond risk premia over the business cycle, for our robustness analysis, it is important to choose out-of-sample periods in which inflation has high information content for bond excess returns. Specifically, we take into consideration both – periods of quite high as well as times when inflation is particularly low. While all inflation models perform about the same during periods of high inflation, we find that in times when inflation is quite low, our iterated Kalman technique significantly dominates (in terms out–of–sample predictability) over the FA and PCA. Hence,we confirm once again that the iterated Kalman technique enables us to extract important inflation information over and above the one that is captured by other computationally simple approaches. At this point we want to stress that in all subsamples we looked at, our iterated Kalman inflation volatility factor has never been significantly outperformed.[9]

We also test the performance of the two filtering techniques in a more challenging framework. Without making any additional assumptions, we create a pool of predictors, including the two different pairs of inflation measures and the CP factor, and let the data

[9] Results based on SPA test. Details available upon request.

themselves choose the most informative variables. This is achieved by finding for each possible number of predictors the subset of the corresponding size that gives the smallest residual sum of squares.[10] Then, we use the Bayesian Schwarz Information Criterion (BIC) to select the best model. We find that regressing the excess returns on the CP factor and the volatility of inflation obtained by the iterated Kalman filter i.e. Model $\mathcal{M}_3$ leads to optimal results.

Finally, we discuss the overall impact of the individual inflation factors in forecasting bond risk premia. Based on the in–sample fit, out–of–sample forecasting, and economic significance, we document that the most important macroeconomic variable for bond excess returns represents the volatility of inflation estimated via the iterated Kalman filter technique. Yet, our inflation volatility measure is no longer a statistically significant predictor of long–term bond risk premia once the level of inflation is in the same regression. The reason for this is the high correlation between the two iterated Kalman filter factors. However, their impact varies with the time to maturity of a bond. In general, we may conclude that the iterated Kalman filter technique allows us to extract in a more accurate way the investors' perceptions of inflation risk in comparison with alternative approaches.

# 4 Conclusions

In this paper we propose a new, computationally simple approach for reconstructing the level and volatility dynamics of a latent macroeconomic factor from a large panel of macroe-

[10]This procedure is known in the literature as best subset selection. See Hastie et al. (2001) for more details.

conomic indices. Our estimation procedure is based on the iterated Kalman filter technique in which we iterate between filtering the unobservable factor with a Kalman filter in the cross–sectional dimension and estimating its variance dynamics in the time series dimension.

We assess the performance of our iterated Kalman filter approach on a set of empirical studies. Extensive simulation results reveal the accuracy of our latent factor volatility estimates and its superiority in comparison with other alternative approaches. Encouraged by those results, we test the ability of our approach to reconstruct in a more accurate way the unobservable macroeconomic driver and its volatility on a real data application – bond risk premia forecasting. Using a panel of a large number of inflation time series, we filter the level and the volatility of inflation via different techniques. We find that in predicting long–term bond risk premia, our inflation estimates significantly outperform the other competitors. In addition, looking at the correspondence between NBER business cycles and inflation fundamentals, we conclude that our estimates are not only statistically but also economically significant.

Our analysis could be taken a step further by studying the performance of bond risk premia in a term structure modeling framework. The iterated Kalman technique could also be to used obtain more accurate estimates for other important macroeconomic predictors such as real activity. However, those extensions are left for future research.

# References


Ang, A. and Piazzesi, M.: 2003, A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* **50**(4), 745–787.

Baillie, R., Bollerslev, T. and Mikkelsen, H.: 1996, Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **74**, 3–30.

Cochrane, J. and Piazzesi, M.: 2005, Bond risk premia, *American Economic Review* **95**, 138–60.

Diebold, F. and Nerlove, M.: 1989, The dynamics of exchange rate volatility: a multivariate latent factor arch model, *Journal of Applied Econometrics* **4**(1), 1–21.

Duffee, G.: 2009, Information in (and not in) the term structure. Working Paper, Haas School of Business, University of California-Berkeley.

Fiorentini, G., Sentana, E. and Shephard, N.: 2004, Likelihood-based estimation of latent generalized arch structures, *Econometrica* pp. 1481–1517.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L.: 2004, The generalized dynamic factor model consistency and rates, *Journal of Econometrics* **119**(2), 231–255.

Forni, M. and Reichlin, L.: 1998, Lets get real: A factor analytic approach to disaggregated business cycle dynamics, *Review of Economic Studies* **65**, 453–473.

Gagliardini, P. and Gourieroux, C.: 2009, Efficiency in large dynamic panel models with common factor. Working Paper Series.

Gürkaynak, R., Sack, B. and Wright, J.: 2006, The u.s. treasury yield curve: 1961 to the present, *Working paper*.

Hansen, P., Lunde, A. and Nason, J.: 2011, The model confidence set, *Econometrica* **79(2)**, 453–497, 03.

Hansen, P. R.: 2005, A test for superior predictive ability, *Journal of Business and Economic Statistics* **23**, 365–380.

Harvey, A., Ruiz, E. and Sentana, E.: 1992, Unobserved component time series models with arch disturbances, *Journal of Econometrics* **52**(1-2), 129–157.

Hastie, T., Tibshirani, R. and Friedman, J.: 2001, *The elements of statistical learning: data mining, inference and prediction*, Springer Series in Statistics, Springer, Canada.

Joslin, S., Priebsch, M. and Singleton, K.: 2009, Risk premium accounting in gaussian macro-dynamic term structure models. Working Paper, MIT.

Ludvigson, S. and Ng, S.: 2009a, A factor analysis of bond risk premia, *Handbook of Applied Econometrics* p. forthcoming.

Ludvigson, S. and Ng, S.: 2009b, Marco factors in bond risk premia, *Review of Financial Studies* **22**(12), 5027–5067.

Sentana, E.: 1991, Time-varying volatility and returns on ordinary shares: an empirical investigation. Unpublished Doctoral Dissertation, University of London.

Sentana, E.: 2004, Factor representing portfolios in large asset markets, *Journal of Econometrics* **119**(2), 257–289.

Sentana, E., Calzolari, G. and Fiorentini, G.: 2008, Indirect estimation of large conditionally heteroskedastic factor models, with an application to the dow 30 stocks, *Journal of Econometrics* **146**(1), 10–25.

Stock, J. and Watson, M.: 2005, Implications of dynamic factor models for var analysis. Working Paper, Princeton University.

Wright, J.: 2009, Term premiums and inflation uncertainty: Empirical evidence from an international panel dataset. Working Paper Series.

# A Data Appendix

This appendix presents U.S. inflation data used in our real data analysis and our estimates for the factor loadings and the parameters of the dynamics of the factor variance.

Table 10 report our inflation data for U.S.: the first column lists the short name of the inflation variable, followed by its mnemonic in column 2, and a brief data description in column 4. All data series are from Global Insights Basic Economic Database. The third column shows the transformations used to assure stationarity of the individual time series. In particular, $\Delta \ln$ and lv denote the first difference of the logarithm and the level of the series, respectively. These data span the period January 1959 - December 2007 for a total of 588 monthly observations.

[Table 5 about here.]

The estimated factor loadings of the inflation indices (in percentage) are:

$\hat{\beta} = [4.77, 5.36, 6.04, 4.53, 6.22, 4.80, 2.72, 5.16, 5.75, 4.38, 3.48, 5.37, 4.82, 4.58, 4.74, 4.23, 2.60, 4.37, 4.57, 4.65]'$, while the estimated constant, ARCH and GARCH parameters of the variance equation of the factor are respectively: 0.0160, 0.1033 and 0.8847.

## B Regression Results

The aim of this Appendix is to provide additional results for assessment of the different pairs of inflation factors (i.e. level and volatility) as predictors for bond excess returns. In particular, we take into consideration the following eight regressions:

$$\begin{aligned}
&\text{Model } \mathcal{M}_1: && rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 \pi_t^{IK} + \varepsilon_{t+12}^{(n)}\\
&\text{Model } \mathcal{M}_2: && rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 \mathtt{vol}\pi_t^{IK} + \varepsilon_{t+12}^{(n)}\\
&\text{Model } \mathcal{M}_3: && rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 \pi_t^{FA} + \varepsilon_{t+12}^{(n)}\\
&\text{Model } \mathcal{M}_4: && rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 \mathtt{vol}\pi_t^{FA} + \varepsilon_{t+12}^{(n)}\\
&\text{Model } \mathcal{M}_5: && rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 \pi_t^{IK} + \gamma_2 \mathtt{vol}\pi_t^{IK} + \varepsilon_{t+12}^{(n)}\\
&\text{Model } \mathcal{M}_6: && rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 \pi_t^{FA} + \gamma_2 \mathtt{vol}\pi_t^{FA} + \varepsilon_{t+12}^{(n)},\\
&\text{Model } \mathcal{M}_7: && rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 \pi_t^{IK} + \gamma_2 \pi_t^{FA} + \varepsilon_{t+12}^{(n)}\\
&\text{Model } \mathcal{M}_8: && rx_{t+12}^{(n)} = \gamma_0 + \gamma_1 \mathtt{vol}\pi_t^{IK} + \gamma_2 \mathtt{vol}\pi_t^{FA} + \varepsilon_{t+12}^{(n)},
\end{aligned}$$

where $rx_{t+12}^{(n)}$ are the excess returns on nominal bond at time $t+12$ with time to maturity $n$ years. The inflation level $\pi_t$ and volatility factors $\mathtt{vol}\pi_t$, estimated by the two different approaches are named as follows: iterated Kalman filter (denoted by $\pi_t^{IK}$ and $\mathtt{vol}\pi_t^{IK}$) and factor analysis (denoted by $\pi_t^{FA}$ and $\mathtt{vol}\pi_t^{FA}$), respectively. Table 4 and Table 5 provide the results.

| | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_6$ | $\mathcal{M}_7$ | $\mathcal{M}_8$ |
|---|---|---|---|---|---|---|---|---|
| Intercept | 0.0294 | 0.0193 | 0.0305 | 0.0314 | 0.0100 | 0.0281 | 0.0167 | 0.0212 |
| | (2.2371) | (2.3795) | (2.5789) | (2.5367) | (0.6550) | (2.3195) | (1.3340) | (1.5528) |
| Inflation Level | -0.0217 | | | | 0.0157 | | 0.0904 | |
| (Iterated Kalman) | (-1.4696) | | | | (0.6277) | | (1.7435) | |
| Inflation Vol | | -0.0076 | | | -0.0114 | | | -0.0067 |
| (Iterated Kalman) | | (-2.0533) | | | (-1.6469) | | | (-1.0244) |
| Inflation Level | | | -0.0406 | | | -0.0811 | -0.1718 | |
| (FA) | | | (-1.9596) | | | (0.0438) | (-2.2289) | |
| Inflation Vol | | | | -0.0409 | | 0.0438 | | -0.0058 |
| (FA) | | | | (-1.8503) | | (0.0942) | | (-0.1698) |
| $R^2$ | 0.051 | 0.086 | 0.081 | 0.072 | 0.092 | 0.083 | 0.127 | 0.086 |
| Adjusted $R^2$ | 0.049 | 0.084 | 0.079 | 0.07 | 0.088 | 0.08 | 0.124 | 0.083 |

Panel B: Predictive Regression Analysis 10 Year Excess Returns

| | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_6$ | $\mathcal{M}_7$ | $\mathcal{M}_8$ |
|---|---|---|---|---|---|---|---|---|
| Intercept | 0.0511 | 0.0294 | 0.0515 | 0.0543 | 0.01408 | 0.05007 | 0.0301 | 0.0342 |
| | ( 2.1400) | (2.0393) | (2.4165) | (2.4120) | (0.5314) | 2.2161 | (1.2913) | |
| Inflation Level | -0.0455 | | | | 0.0259 | | 0.1403 | |
| (Iterated Kalman) | (-1.6964) | | | | (0.6057) | | (1.5030) | |
| Inflation Vol | | -0.0156 | | | -0.0218 | | | -0.0134 |
| (Iterated Kalman) | | (-2.3815) | | | (-1.8924) | | | (-1.2325) |
| Inflation Level | | | -0.0810 | | | -0.1063 | -0.2845 | |
| (FA) | | | (-2.1616) | | | (-0.6660) | (-2.0646) | |
| Inflation Vol | | | | -0.0836 | | 0.0274 | | -0.0142 |
| (FA) | | | | (-2.0871) | | (0.1612) | | (-0.2410) |
| $R^2$ | 0.063 | 0.100 | 0.090 | 0.085 | 0.105 | 0.091 | 0.122 | 0.101 |
| Adjusted $R^2$ | 0.062 | 0.099 | 0.089 | 0.084 | 0.101 | 0.087 | 0.119 | 0.097 |

Table 4: Results for ordinary least squares regressions for eight different models (labeled as $\mathcal{M}_1$, $\mathcal{M}_2$,. . .,$\mathcal{M}_8$) utilizing annual returns on 5- and 10-year Treasury bonds. Standard errors are corrected for autocorrelation and heteroskedasticity. t-statistics are reported in parenthesis. Asterisks $^*$ ,$^{**}$ ,$^{***}$ indicate statistical significance at the 10%, 5%, and 1% level, respectively. The data span the period June 1962 to December 2003. See text

| | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_6$ | $\mathcal{M}_7$ | $\mathcal{M}_8$ |
|---|---|---|---|---|---|---|---|---|
| Intercept | 0.0819 | 0.0393 | 0.0827 | 0.0894 | 0.0175 | 0.0843 | 0.0421 | 0.0595 |
| | (1.9551) | (1.5634) | (2.2193) | (2.2349) | (0.4102) | (2.0606) | (1.0390) | (1.3660) |
| Inflation Level | -0.0875 | | | | 0.03676 | | 0.2660 | |
| (Iterated Kalman) | (-1.7849) | | | | (0.5252) | | (1.6159) | |
| Inflation Vol | | -0.0291 | | | -0.0380 | | | -0.0199 |
| (Iterated Kalman) | | (-2.5615) | | | (-2.3542) | | | (-1.3236) |
| Inflation Level | | | -0.1554 | | | -0.1280 | -0.5413 | |
| (FA) | | | (-2.2750) | | | (-0.5619) | (-2.1927) | |
| Inflation Vol | | | | -0.1634 | | -0.0296 | | -0.0602 |
| (FA) | | | | (-2.2194) | | (-0.1190) | | (-0.6041) |
| $R^2$ | 0.065 | 0.097 | 0.092 | 0.090 | 0.100 | 0.093 | 0.124 | 0.100 |
| Adjusted $R^2$ | 0.063 | 0.096 | 0.091 | 0.089 | 0.096 | 0.089 | 0.121 | 0.096 |

Panel B: Predictive Regression Analysis: 30 Year Excess Returns

| | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_6$ | $\mathcal{M}_7$ | $\mathcal{M}_8$ |
|---|---|---|---|---|---|---|---|---|
| Intercept | 0.1245 | 0.05139 | 0.1255 | 0.1382 | 0.0427 | 0.1333 | 0.0598 | 0.1133 |
| | (1.9933) | (1.3845) | (2.2087) | (2.2573) | (0.6766) | (2.1394) | (1.1065) | (1.6143) |
| Inflation Level | -0.1431 | | | | 0.0145 | | 0.4303 | |
| (Iterated Kalman) | (-1.9879) | | | | (0.1466) | | (1.8445) | |
| Inflation Vol | | -0.0447 | | | -0.0483 | | | -0.0165 |
| (Iterated Kalman) | | (-2.5900) | | | (-2.2256) | | | -0.8505 |
| Inflation Level | | | -0.2539 | | | -0.1216 | -0.8780 | |
| (FA) | | | (-2.4453) | | | (-0.4194) | (-2.3525) | |
| Inflation Vol | | | | -0.2701 | | -0.1431 | | -0.1844 |
| (FA) | | | | (-2.4300) | | (-0.4558) | | (-1.2050) |
| $R^2$ | 0.065 | 0.085 | 0.092 | 0.092 | 0.085 | 0.092 | 0.122 | 0.094 |
| Adjusted $R^2$ | 0.063 | 0.083 | 0.090 | 0.090 | 0.082 | 0.089 | 0.119 | 0.090 |

Table 5: Results for ordinary least squares regressions for eight different models (labeled as $\mathcal{M}_1$, $\mathcal{M}_2$,...,$\mathcal{M}_8$) utilizing annual returns on 20- and 30-year Treasury bonds. Standard errors are corrected for autocorrelation and heteroskedasticity. t-statistics are reported in parenthesis. Asterisks $^{*}$ ,$^{**}$ ,$^{***}$ indicate statistical significance at the 10%, 5%, and 1% level, respectively. The data span the period June 1962 to December 2003. See text

| | Simple Average | cross-section OLS | Factor Analysis | Principal Component | Iterated Kalman |
|---|---|---|---|---|---|
| | | | Panel A | | |
| Avg correlation on $f_t$ | 0.9640 | 0.9526 | 0.9898 | 0.9433 | 0.9899 |
| Avg RMSE on $f_t$ | 0.2742 | 0.3161 | 0.1467 | 0.3337 | 0.1391 |
| Avg correlation on $\sigma_t$ | 0.9477 | 0.9379 | 0.9677 | 0.9289 | 0.9691 |
| Avg RMSE on $\sigma_t$ | 0.1404 | 0.1456 | 0.1293 | 0.1460 | 0.0548 |

Table 6: Performance comparison of different filtering methods for the factor dynamics and its conditional volatility over 50 years and 1000 simulation paths. The methods are: simple cross–sectional averages, cross–sectional OLS regression, Factor Analysis, Principal Component, and Iterated Kalman filter. The performance measures are the average correlation and the average Root Mean Square Error (RMSE).

## Performance Comparison - Simulations

| | Simple Average | cross-section OLS | Factor Analysis | Principal Component | Iterated Kalman |
|---|---|---|---|---|---|
| Panel A | | | | | |
| Avg correlation on $f_t$ | 0.9636 | 0.9517 | 0.9900 | 0.9416 | 0.9902 |
| Avg RMSE on $f_t$ | 0.2756 | 0.3193 | 0.1455 | 0.3382 | 0.1381 |
| Avg correlation on $\sigma_t$ | 0.9454 | 0.9340 | 0.9670 | 0.9244 | 0.9684 |
| Avg RMSE on $\sigma_t$ | 0.1397 | 0.1449 | 0.1285 | 0.1451 | 0.0549 |
| Panel B | | | | | |
| Avg correlation on $f_t$ | 0.9635 | 0.9521 | 0.9895 | 0.9426 | 0.9899 |
| Avg RMSE on $f_t$ | 0.2757 | 0.3181 | 0.1468 | 0.3361 | 0.1396 |
| Avg correlation on $\sigma_t$ | 0.6176 | 0.6077 | 0.6397 | 0.5993 | 0.6444 |
| Avg RMSE on $\sigma_t$ | 0.2260 | 0.2291 | 0.2195 | 0.2271 | 0.2032 |

Table 7: Performance comparison of different filtering methods for the factor dynamics and its conditional volatility over 50 years and 1000 simulation paths. The methods are: simple cross–sectional averages, cross–sectional OLS regression, Factor Analysis, Principal Component, and Iterated Kalman filter. The performance measures are the average correlation and the average Root Mean Square Error (RMSE). The different DGP in the two panels are given by the one factor model with: GARCH dynamics in $\Delta_t$ and $\Phi_t$ (Panel A), GARCH dynamics with two regimes in $\Delta_t$ (Panel B).

## Performance Comparison - Simulations

| | Simple Average | cross-section OLS | Factor Analysis | Principal Component | Iterated Kalman |
|---|---|---|---|---|---|
| **Panel A** | | | | | |
| Avg correlation on $f_t$ | 0.9633 | 0.9513 | 0.9898 | 0.9414 | 0.9899 |
| Avg RMSE on $f_t$ | 0.2769 | 0.3205 | 0.1499 | 0.3406 | 0.1395 |
| Avg correlation on $\sigma_t$ | 0.6019 | 0.5907 | 0.6255 | 0.5810 | 0.6307 |
| Avg RMSE on $\sigma_t$ | 0.2229 | 0.2260 | 0.2166 | 0.2240 | 0.2032 |
| **Panel B** | | | | | |
| Avg correlation on $f_t$ | 0.9636 | 0.9520 | 0.9897 | 0.9422 | 0.9899 |
| Avg RMSE on $f_t$ | 0.2757 | 0.3180 | 0.1579 | 0.3419 | 0.1394 |
| Avg correlation on $\sigma_t$ | 0.5609 | 0.5486 | 0.5905 | 0.5393 | 0.5936 |
| Avg RMSE on $\sigma_t$ | 0.2115 | 0.2146 | 0.2054 | 0.2125 | 0.1988 |

Table 8: Performance comparison of different filtering methods for the factor dynamics and its conditional volatility over 30 (Panel A) and 15 years (Panel B). The methods are: simple cross–sectional averages, cross–sectional OLS regression, Factor Analysis, Principal Component, and Iterated Kalman filter. The performance measures are the average correlation and the average Root Mean Square Error (RMSE).

Panel A: Out-of-Sample Mean Squared Errors

| | 5Y Bond Exret | 10Y Bond Exret | 20Y Bond Exret | 30Y Bond Exret |
|---|---|---|---|---|
| Model $\mathcal{M}_1$ | 0.0033 (0.0679) | 0.0091 (0.0163) | 0.0320 (0.0000) | 0.0873 (0.0000) |
| Model $\mathcal{M}_2$ | 0.0029 (0.3114) | 0.0074 (0.1292) | 0.0256 (0.0818) | 0.0706 (0.0860) |
| Model $\mathcal{M}_3$ | 0.0028 (0.7174) | 0.0070 (0.4381) | 0.0241 (0.4289) | 0.0674 (0.3816) |
| Model $\mathcal{M}_4$ | 0.0029 (0.1424) | 0.0076 (0.0060) | 0.0265 (0.0007) | 0.0731 (0.0000) |
| Model $\mathcal{M}_5$ | 0.0029 (0.3086) | 0.0075 (0.0067) | 0.0259 (0.0006) | 0.0709 (0.0090) |
| Model $\mathcal{M}_6$ | 0.0028 (0.5228) | 0.0070 (0.6444) | 0.0242 (0.5927) | 0.0683 (0.6727) |
| Model $\mathcal{M}_7$ | 0.0030 (0.3163) | 0.0076 (0.0265) | 0.0260 (0.0007) | 0.0707 (0.0186) |

Panel B: Out–of–Sample Mean Absolute Errors

| | 5Y Bond Exret | 10Y Bond Exret | 20Y Bond Exret | 30Y Bond Exret |
|---|---|---|---|---|
| Model $\mathcal{M}_1$ | 0.0418 (0.0954) | 0.0786 (0.0000) | 0.1576 (0.0000) | 0.2473 (0.0000) |
| Model $\mathcal{M}_2$ | 0.0400 (0.4728) | 0.0697 (0.1435) | 0.1381 (0.0681) | 0.2146 (0.1380) |
| Model $\mathcal{M}_3$ | 0.0391 (0.7175) | 0.0676 (0.3838) | 0.1349 (0.4187) | 0.2120 (0.4242) |
| Model $\mathcal{M}_4$ | 0.0439 (0.1362) | 0.0709 (0.015) | 0.1412 (0.0006) | 0.2200 (0.0000) |
| Model $\mathcal{M}_5$ | 0.0422 (0.2092) | 0.0704 (0.0257) | 0.1396 (0.0010) | 0.2169 (0.0018) |
| Model $\mathcal{M}_6$ | 0.0390 (0.5643) | 0.0676 (0.6722) | 0.1348 (0.6992) | 0.2118 (0.7385) |
| Model $\mathcal{M}_7$ | 0.0440 (0.4819) | 0.0709 (0.0066) | 0.1398 (0.0030) | 0.2165 (0.0016) |

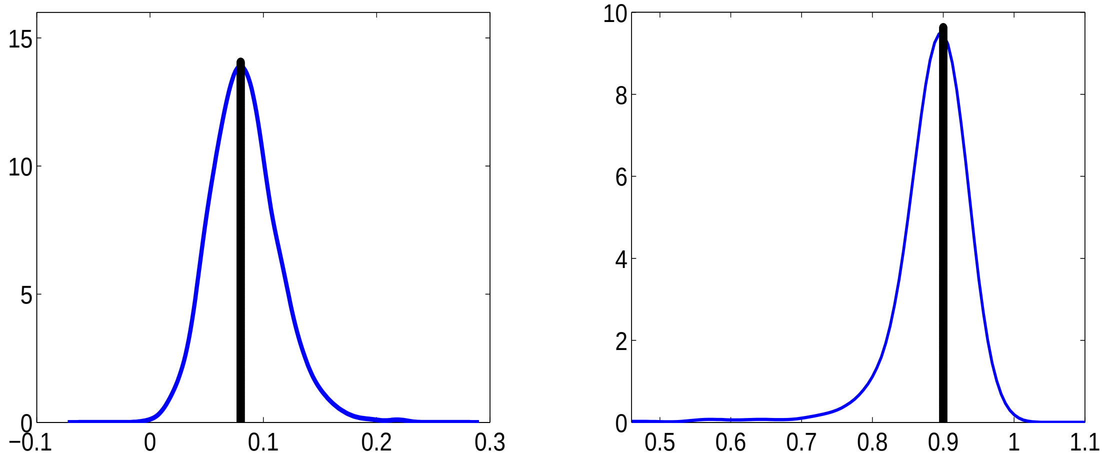
Table 9: Results (mean squared errors (Panel A) and mean absolute errors (Panel B)) of out–of–sample forecasting performance of seven different models for 5-, 10-, 20- and 30-year Treasury Bond excess returns, as described in detail in the text. p-values of the superior predictive ability (SPA) test of Hansen (2005) are reported in parenthesis. The results are based on the out-of-sample period, January 2004 - December 2008.

| Short Name | Mnemonic | Tran | Description |
|---|---|---|---|
| PPI: fin gds | pwfsa | $\Delta$ ln | Producer Price Index: Finished Goods (82=100,Sa) |
| PPI: cons gds | pwfcsa | $\Delta$ ln | Producer Price Index: Finished Consumer Goods (82=100,Sa) |
| PPI: int materials | pwimsa | $\Delta$ ln | Producer Price Index:Intermed Mat.Supplies & Components(82==100,Sa) |
| PPI: crude matls | pwcmsa | $\Delta$ ln | Producer Price Index: Crude Materials (82=100,Sa) |
| Spot market price | psccom | $\Delta$ ln | Spot market price index: bls & crb: all commodities(1967=100) |
| PPI: nonferrous materials | pw102 | $\Delta$ ln | Producer Price Index: Nonferrous Materials (1982=100, Nsa) |
| NAPM com price | pmcp | lv | Napm Commodity Prices Index (Percent) |
| CPI-U: all | punew | $\Delta$ ln | Cpi-U: All Items (82-84=100,Sa) |
| CPI-U: apparel | pu83 | $\Delta$ ln | Cpi-U: Apparel & Upkeep (82-84=100,Sa) |
| CPI-U: transp | pu84 | $\Delta$ ln | Cpi-U: Transportation (82-84=100,Sa) |
| CPI-U: medical | pu85 | $\Delta$ ln | Cpi-U: Medical Care (82-84=100,Sa) |
| CPI-U: comm. | puc | $\Delta$ ln | Cpi-U: Commodities (82-84=100,Sa) |
| CPI-U: dbles | pucd | $\Delta$ ln | Cpi-U: Durables (82-84=100,Sa) |
| CPI-U: services | pus | $\Delta$ ln | Cpi-U: Services (82-84=100,Sa) |
| CPI-U: ex food | puxf | $\Delta$ ln | Cpi-U: All Items Less Food (82-84=100,Sa) |
| CPI-U: ex shelter | puxhs | $\Delta$ ln | Cpi-U: All Items Less Shelter (82-84=100,Sa) |
| CPI-U: ex med | puxm | $\Delta$ ln | Cpi-U: All Items Less Medical Care (82-84=100,Sa) |
| PCE defl | gmdc | $\Delta$ ln | Pce, Impl Pr Defl:Pce (2000=100) (AC) (BEA) |
| PCE defl: dlbes | gmdcd | $\Delta$ ln | Pce, Impl Pr Defl:Pce; Durables (2000=100) (AC) (BEA) |
| PCE defl: nondble | gmdcn | $\Delta$ ln | Pce, Impl Pr Defl:Pce; Nondurables (2000=100) (AC) (BEA) |
| PCE defl: service | gmdcs | $\Delta$ ln | Pce, Impl Pr Defl:Pce; Services (2000=100) (AC) (BEA) |

Table 10: U.S. inflation data. Columns: name of the inflation variable, mnemonic, type of transformation, and data description.

Panel A:

Estimates Pdf of $\alpha = 0.08$ | Estimates Pdf of $\beta = 0.90$




Panel B:

Estimates Pdf of $\alpha = 0.08$ | Estimates Pdf of $\beta = 0.90$

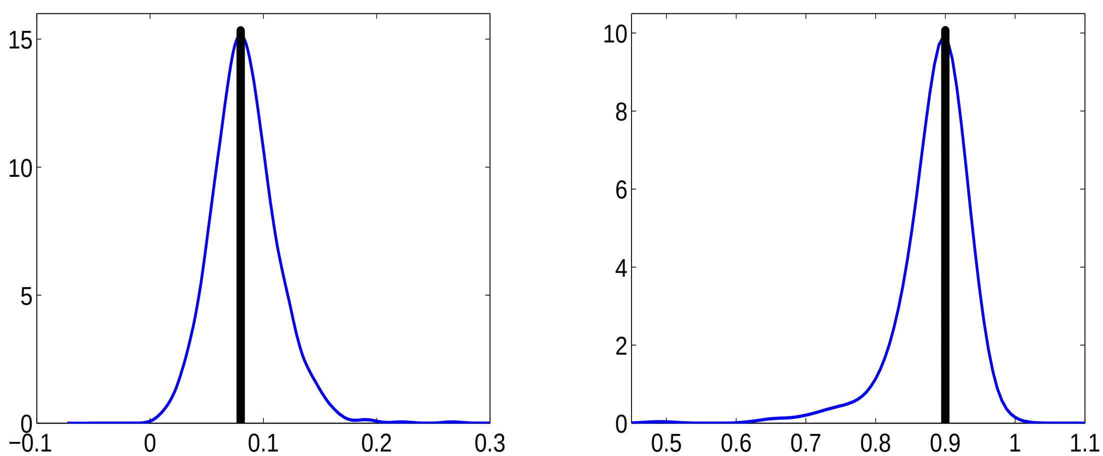
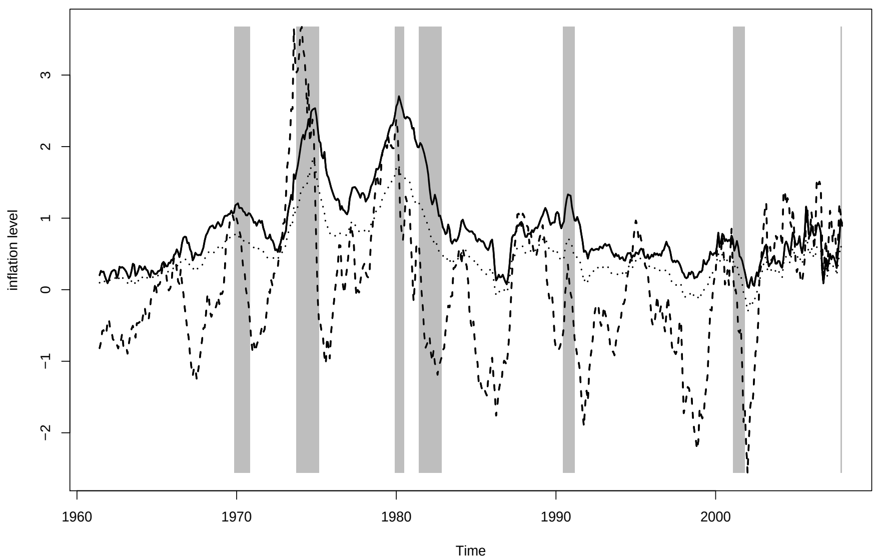



Figure 1: Probability distribution function of the estimation error over 1000 simulation paths of the parameters of the GARCH(1,1) process for the factor conditional variance $\sigma_t^2 = c + \alpha f_{t-1}^2 + \beta\sigma_{t-1}^2$ in a DGP with constant (Panel A) and time varying (GARCH type, Panel B) idiosyncratic noise $\Phi_t$.

Inflation Level




Inflation Volatility

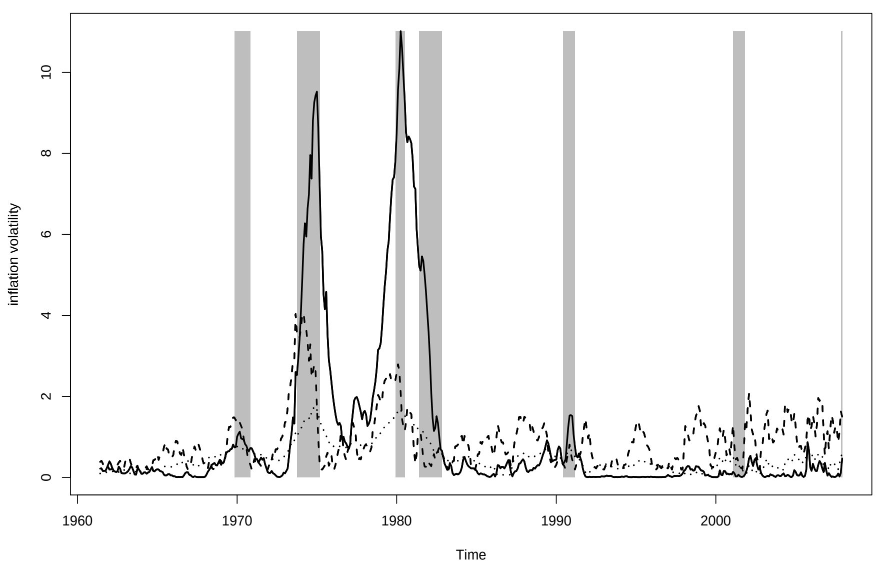



Figure 2: The upper panel plots the three estimates of inflation level: iterated Kalman filter (solid line), factor analysis (dotted line), and principal components (dashed line) based on a panel of 21 inflation time series, as described in the text. The lower panel plots the inflation volatility filtered by the three techniques. Once again the solid line indicates the iterated Kalman filter estimate, the dotted line the estimate got applying factor analysis, whereas the dashed line represents the dynamics of the principal components volatility. The shaded bars denote months designated as recessions by NBER.