Specification Analysis of International Treasury Yield Curve Factors

Fulvio Pegoraro (Banque de France, CREST and HEC Lausanne)

Andrew F. Siegel (University of Washington Business School)

Luca Tiozzo 'Pezzoli' (Banque de France and Paris Dauphine University)

March 2012

Extended Abstract:

We address theoretical and empirical issues in the estimation of interest rate factors that evolve over time in a multi country setting. The yield curve literature, following the seminal papers of Vasicek (1977) and Cox, Ingerssoll and Ross (1983), has focused not only on the specification and estimation of models explaining the term structure of interest rates in a single economy but, more recently, has also tried to handle the relevant problem of specifying and estimating the joint dynamics of international yield curves.

In the single-country case, empirical studies have found that two or three latent factors [the so called level, slope and curvature factors of Litterman and Scheinkman (1991)] are required by a yield curve model to match the dynamics and the shapes of the term structure, and this is regardless of the sample period and the data source [see Dai and Singleton (2000, 2002, 2003), Duffee (2002), Cheridito, Filipovic and Kimmel (2002), Duarte (2004)]. This wide degree of robustness has made this result a well known fundamental building block characterizing the modeling of single-country yield curves.

In the multi-country setting, on the contrary, we observe a relevant lack of agreement, not only about the number of latent factors that are required to explain the joint dynamics of two or more yield curves, but also about their nature, namely how many are common factors (i.e., they affect yields in all countries) and how many are local factors (i.e. they affect yields in one country only) [see, among the others, Backus, Foresi and Telmer (2001), Han and Hammond (2003), Ahn (2004), Leippold and Wu (2007), Diebold, Li and Yue (2008) and Egorov, Li and Ng (2011)].

The purpose of this paper is to provide, by means of a state-space approach and working with a new international Treasury zero-coupon bond database, a reliable detection of the optimal numbers of common and local factors required

to jointly explain multi-country yield curves, and to obtain the optimal (in the conditional mean square error sense) Kalman-based extraction of these latent factors in order to more precisely understand the linkages between international bond markets and macro variables.

We focus on two possible explanations for these contrasting conclusions about the number of latent factors and their nature (common or local). First, these contrasting conclusions may depend on the fact that the empirical analyses are run for different numbers of countries or for different sample periods. Second, this lack of agreement may also depend on the statistical techniques adopted to determine the relevant number of latent factors and to identify them as common or local ones.

To address the first issue, we build an international Treasury yield curve database at daily frequency, from the beginning of the 1980s to the end of 2009, for four leading bond markets: the U.S., Germany, U.K. and Japan. We adopt for all countries the Gurkaynack, Sack and Wright (2007) criteria to filter coupon bond Treasury raw data (taken from Datastream), and to guarantee a uniform level of liquidity. Indeed, Gurkaynack, Sack and Wright (2007) exclude not only bonds with any option-like features and those with too short or too long maturities (i.e. not actively traded), but also prevent the possible presence of an on-the-run/off-the-run liquidity premium because on-the-run and first off-the-run bonds have been filtered out. We then interpolate the discount function, across residual maturities and currencies, using the (parsimonious smoothed) Nelson and Siegel (1987) methodology. We thus obtain what we believe to be the first international Treasury yield curve data base at daily frequency, with a homogeneous interpolation technique and a uniform level of bond liquidity (of an actively traded second off-the-run kind) across time, maturities and economies. These features clearly facilitate a) reliable comparison (in a way explained below) between different numbers and combinations of common and local yield factors and b) identification of single-country or multi-country macro-financial variables (that may be linked to common or local variables) that may be adopted to help explain international term structure variation across time and maturities.

Regarding the statistical techniques adopted to determine the relevant numbers of common and local factors, the international yield curve literature (as cited above) has typically applied principal component analysis or principal factor analysis. Nevertheless, two main critiques about this choice have appeared in the literature. First, Pérignon, Smith and Villa (2007) state that the purpose of principal component analysis is to extract factors that maximize the explained variance, and do not seek to distinguish between the role of common and that of local factor in the presence of multiple groups. Thus a key drawback of using

principal component analysis in presence of multiple groups (e.g., international term structure of interest rates across countries) is the fact that estimated factors jointly capture both local and common influences [see Section 2.3 of Pérignon, Smith and Villa (2007)] without distinguishing one from the other. Second, the factor model literature has proposed several methods for selecting the number of common factors. However, reliability of these criteria requires the presence of weak-form serial and cross-sectional dependence in the idiosyncratic component of the factor model, as well as large N (the cross-sectional dimension) and T (the time-series dimension) in the database of interest. These conditions are clearly not all satisfied by an international yield curve panel of data, given the strong persistence and cross-correlation of interest rates, as well as the typically small dimension of the maturity spectrum. For instance, in presence of serial dependence, the Bai and Ng (2002) criteria tend to overestimate the number of common factors with considerable probability, and this problem stands out even when a first-difference filter is applied to stationary data in order to mitigate the persistence [see Greenaway-McGrevy, Han and Sul (2011) for details]. In addition, when N is small the estimator of the factor score is not consistent, that is, the extraction of the latent factor is potentially not reliable.

In order to detect the optimal numbers of common and/or local factors, we adopt a linear Gaussian state-space approach in which the joint dynamics of multi-country term structures is explained by autoregressive stationary latent factors that are identified as common or local by means of a proper parameterization of the measurement equation. More precisely, a factor is defined as common if the associated factor loadings are permitted to be nonzero for all economies, while it is local to a given country if its factor loadings are zero for all other countries. Next, for any given set of international yield curves, we use maximum likelihood estimation for the state-space model, based on the EM algorithm and Kalman filter and smoother recursions [see Engle and Watson (1981), Quah and Sargent (2004) and Doz, Giannone and Reichlin (2006)]. From among different scenarios, each specifying the numbers and combinations of common and local factors, we select the optimal combination on the basis of maximum likelihood-based criteria and we provide a comparison with selection methodology suggested by recently followed principal component-based approaches [see Leippold and Wu (2007), Diebold, Li and Yue (2008), and Egorov, Li and Ng (2011)].

The algorithm is fast and powerful. First, its complexity depends on the (quite small) number of latent factors and not on the (much larger) dimension of the cross-section; second, at each iteration of the EM algorithm the likelihood increases and (under regularity conditions) it converges to the maximum of the likelihood function; third, the number of iterations that are necessary to reach convergence is in general small given that the algorithm is initialized with

principal-component-based estimates. Our approach has the advantage, as compared with the principal component or principal factor approaches, to be capable of identifying the optimal number of common and/or local factors, as well as optimally extracting these dynamic latent factors, in a way that is compatible with panel data with small N, large T, and strong dependence (both serial and cross-sectional). Finally, we study the links between the optimally extracted factors and macro-financial variables (both single-country and multi-country) and provide a comparison to the literature.