Model Selection with Estimated Factors and Idiosyncratic Components

Jack Fosten[*]

University of East Anglia, UK

October 5, 2016

Abstract

This paper provides consistent information criteria for the selection of forecasting models which use a subset of both the idiosyncratic and common factor components of a big dataset. This hybrid model approach has been explored by recent empirical studies to relax the strictness of pure factor-augmented model approximations, but no formal model selection procedures have been developed. The main difference to previous factor-augmented model selection procedures is that we must account for estimation error in the idiosyncratic component as well as the factors. Our main contribution is to show the conditions required for selection consistency of a class of information criteria which reflect this additional source of estimation error. We show that existing factor-augmented model selection criteria are inconsistent in circumstances where N is of larger order than $\sqrt{T}$, where N and T are the cross-section and time series dimensions of the dataset respectively, and that the standard BIC is inconsistent regardless of the relationship between N and T. We therefore propose a new set of information criteria which guarantee selection consistency in the presence of estimated idiosyncratic components. The properties of these new criteria are explored through a Monte Carlo simulation study. The paper concludes with an empirical application to long-horizon exchange rate forecasting using a recently proposed model with country-specific idiosyncratic components from a panel of global exchange rates.

JEL Classification: C13, C22, C38, C52, C53

Keywords: Forecasting, Factor model, model selection, information criteria, idiosyncratic

1 Introduction

This paper provides consistent model selection criteria in predictive regressions involving both the common factors and the idiosyncratic components from a big dataset. The modelling environment differs from the standard two-stage "diffusion index" approach of Stock and Watson (2002a,b), where only the estimated factors are retained from the first-stage to be used as regressors in the

1

second-stage forecasting model, and all remaining idiosyncratic information from the big dataset is discarded. We argue that the pure factor approach may be an excessive approximation in cases where a target variable has a strong relationship with a particular set of variables in the dataset. As an example, we might specify a forecasting model for inflation by exploiting the idiosyncratic variation of a small number unemployment variables as predicted by a Phillips curve model, and also use the factors in order to pick up the 'big data' effect. Hybrid models of this form have been considered as early as Stock and Watson (1999) and can provide benefits in terms of forecast accuracy by performing significant data reduction and limiting the effect of the so-called "curse of dimensionality". Recent empirical studies such as Castle et al. (2013), Luciani (2014) and Engel et al. (2015) have also used this type of hybrid model for macroeconomic forecasting. In the financial econometrics literature, models involving the idiosyncratic component also arise in studies such as Brownlees et al. (2015) which analyse networks in asset returns by using 'de-factored' log-prices obtained by subtracting the common factor component of each asset return.

The objective of this paper is to provide information criteria for model selection in a second-stage forecasting model which additionally uses the estimated idiosyncratic components; something which has not been addressed in the existing literature to the best of our knowledge. There has, however, been recent progress in research into model selection criteria in "pure" factor-augmented models, which do not involve the idiosyncratic components. Bai and Ng (2009) proposed a boosting approach to determine the number of autoregressive lags and factors which enter the forecasting model. Other approaches include Groen and Kapetanios (2013), who propose modified Bayesian and Hannan-Quinn type information criteria and Djogbenou (2016) who uses cross-validation. Earlier research focussed on the first stage of this two-step process, for selecting the number of factors present in the big dataset. The seminal paper of Bai and Ng (2002) showed how to modify standard information criteria when both the cross-section (N) and time series dimension (T) of the dataset grow to infinity. Subsequently, Amengual and Watson (2007), Hallin and Liška (2007) and Onatski (2010) have proposed different methods to tackle the same problem. However, none of these approaches are able to deal with the case where both factors and idiosyncratic components are estimated and used in the forecasting model. This yields new challenges which we must address.

The main issue which is new to this paper relative to the literature on pure factor-augmented models, is that we additionally use estimates of the idiosyncratic components as forecast model regressors. Our main contribution is to propose new information criteria for model selection which take into account the new source of estimation error which results from the estimation of the idiosyncratic component. We use a general class of information criteria with a penalty function $g(N,T)$ which depends both on N and T. The main result is a theorem on selection consistency, which shows that consistency only obtains for information criteria with a penalty satisfying the condition $\min\left\{\sqrt{T}, N\right\} g(N,T) \to \infty$, when Principal Components is used for estimation. This rate condition is a new finding in the literature, and is different to existing criteria for pure factor-augmented models. The result is driven by a Lemma which shows that estimation error in the idiosyncratic components vanishes at a rate $\min\left\{\sqrt{T}, N\right\}$, whereas the estimation error in the

factors vanishes at a rate $\min\{T, N\}$.[1]

This new rate condition carries several important implications. Firstly, it implies that existing factor-augmented model selection procedures are inconsistent in cases where N is large relative to $\sqrt{T}$ for selecting models involving the estimated idiosyncratic components. Since it is quite natural for N to be of similar order to T in macroeconomic forecasting, this inconsistency in not restricted to trivial cases. Secondly, the result means that the standard Bayesian Information Criterion (BIC), commonly used in time series applications, is inconsistent for any relative rate of increase between N and T. These findings show that we must make modifications to standard information criteria in order to get consistent model selection. We therefore propose a range of new criteria using different functional forms whose properties are equivalent asymptotically but vary in finite samples. We investigate these finite sample properties through a set of Monte Carlo experiments which demonstrate the improvements of our methods relative to related criteria such as those of Groen and Kapetanios (2013).

Our paper bears similarities to other econometric literatures of factor-augmented model specification, in addition to the aforementioned literature on pure factor-augmented models. Model selection methods are closely related to model averaging methods such as Hansen (2007) and, for the factor-augmented case, Cheng and Hansen (2015). While these model averaging methods are somewhat more general as they nest the model selection methods used here,[2] they are more complex as they require the estimation of model weights, in addition to the estimation of both the factor model and the factor-augmented forecasting model. The factor model approach we employ also has similarities with the rapidly expanding econometric literature on 'big data' methods more generally. The properties and estimation of competing data reduction methods has been comprehensively studied by De Mol et al. (2008) and more recently Carrasco and Rossi (2016). While a comparison of competing big data methods is not the aim of this study, it is useful to keep in mind that these comparisons would also change when introducing the idiosyncratic components to the forecasting model.

We apply our new model selection criteria to the challenging but important empirical problem of long-horizon exchange rate forecasting. The recent paper of Engel et al. (2015) proposes a factor-based approach to exchange rate forecasting. They suggest to use the idiosyncratic component from a global dataset of countries' exchange rates as the 'fundamental' in a regression model for a particular country's exchange rate. Therefore their approach precisely matches the modelling framework under which our methods apply. We extend their model to allow for cross-country exchange rate spillovers, modelled by the idiosyncratic components, and use our new information criteria to select between these spillover effects. Our results, applied to a range of OECD countries, show that it is very difficult to out-perform a naïve no-change model; a result mirrored in the majority of existing empirical evidence following the seminal work of Meese and Rogoff (1983b,a).

---

[1]The rate for factor estimation error was shown in Theorem 1 of Bai and Ng (2002), Lemmas B.1-B.3 of Bai (2003) and Lemma A.1 of Bai and Ng (2006).

[2]Model selection is equivalent to model averaging, where the selected model receives a weight of one and all other models receive a weight of zero.

However, we also find that our model selection criteria select a non-trivial number of idiosyncratic effects both in-sample and out-of-sample. This is in contrast to the standard BIC which is not consistent and severely overfits the model by selecting the maximum possible number of variables in many cases.

The rest of the paper is organized as follows. Section 2 provides the forecasting set-up and motivates the use of idiosyncratic components alongside the factors. Section 3 introduces the class of information criteria we use, and provides the new asymptotic rates for the penalty function which are required for selection consistency. Section 4 proposes specific functional forms for the penalty function which satisfy the conditions required for consistency, and compares them to existing information criteria. Section 5 provides a Monte Carlo analysis. Section 6 presents the empirical application to exchange rate forecasting and Section 7 concludes the paper.

2 Set-up and Motivation

The broad interest of this study is in predictive regressions for a target variable y_{t+h} when a large set of N candidate predictor variables, X_t, are available. In the first stage of the two-step procedure of Stock and Watson (2002a,b), we assume that the predictors permit the common factor structure:

$$X_t = \Lambda F_t + u_t \tag{1}$$

where F_t is an unobserved $r \times 1$ vector of common factors, Λ is an $N \times r$ matrix of factor loadings and u_t an $N \times 1$ vector of idiosyncratic errors. In the theory part of this paper, in the same way as related papers such as Cheng and Hansen (2015), we will assume that the number of factors, r, is known. In practice, of course, r is not known to the applied researcher and can instead be consistently estimated by the information criteria of Bai and Ng (2002). The unknown factors and loadings can then be estimated by methods such as principal components (PCA) as in Stock and Watson (2002a,b) and Bai and Ng (2002, 2006).[3]

In the second stage, we propose the general predictive regression which has the following form:

$$y_{t+h} = \beta^{0\prime} F_t^0 + \alpha^{0\prime} u_t^0 + \varepsilon_{t+h} \tag{2}$$

where F_t^0 is an $r^0 \times 1$ vector which is a subset of the factors, F_t, so that $r^0 \leq r$. This allows for situations either where y_{t+h} is affected by all r factors, or where only some subset of the factors has predictive ability for y_{t+h}. Similarly, u_t^0 is a finite $m^0 \times 1$ subset of the idiosyncratic error vector u_t. The identity of F_t^0 and u_t^0 is unknown and their selection is the main aim of this study. When $r^0 + m^0 << N$, significant data reduction can improve predictions of y_{t+h} by reducing the excess variability caused by parameter uncertainty.[4]

---

[3]Other methods, such the maximum likelihood approach of Doz et al. (2011, 2012), could also be used to estimate the factors but we focus on PCA in this paper.

[4]We could also specify Equation (2) to include a set of other 'must-have' non-factor regressors, W_t, as in Bai and Ng (2006) but we omit these here for clarity.

Forecasting models involving idiosyncratic components, where $\alpha^0 \neq 0$ in Equation (2), are a relatively recent development in the literature, with an increasing number of applications. Luciani (2014) suggests that the use of additional idiosyncratic terms may be useful in forecasting due to the effect of "pervasive shocks" which affect multiple variables. Castle et al. (2013) choose to look at hybrid models involving factors and variables, and use the *Autometrics* routine to select between alternative formulations. Engel et al. (2015) directly specify long-horizon exchange rate forecasting models as a function of the idiosyncratic component from a factor model of international exchange rates. Studies of financial asset networks such as that of Brownlees et al. (2015), mentioned above, also use a similar structure to Equation (2), although they are not explicitly focussed on the use of these equations in forecasting.

On the other hand, the "pure" factor-augmented approach, where $\alpha^0 = 0$ and only the factors are used in forecasting, has been used extensively in the applied forecasting literature. When $\alpha^0 = 0$ and F_t^0 is equal to the full factor vector F_t with no factors omitted, Equation (2) corresponds exactly to the factor-augmented, or "diffusion index" model of Stock and Watson (2002a,b) and Bai and Ng (2006). This model has been widely used in empirical studies; see Stock and Watson (2011) for an overview. When $\alpha^0 = 0$ but F_t^0 is a smaller subset of F_t, only some factors are used in forecasting a given target variable. This type of model was motivated by Boivin and Ng (2006) in the context of real versus nominal macroeconomic factors. As mentioned above, model selection techniques have been proposed for the pure factor-augmented approach by Bai and Ng (2009) and Groen and Kapetanios (2013), but not for models involving estimated idiosyncratic components.

In addition to the empirical motivation for pursuing the model in Equation (2), we can also offer a simple analytical motivation for using this model. The specification of models involving the idiosyncratic terms has been somewhat overlooked in the literature for the two-stage approach of Stock and Watson (2002a,b). As a starting point, we let y_t be a function of the observable variables X_t, noting that economic theories propose relationships between observed economic variables rather than the latent factors, F_t, which are merely used as a data reduction method and do not have a direct interpretation. In this case, we use the factor model from Equation (1) simply as a way to approximate X_t and obtain a low-dimensional model.

For this analytical example, the data generating process (DGP-X) for y_{t+h} is known to be the linear function of X_t, which we partition into X_t^0 and X_t^1 which are of dimension $m^0 \times 1$ and $(N - m^0) \times 1$ respectively, where m^0 is a finite integer as specified above.

(DGP-X) $$y_{t+h} = \alpha^{0\prime}X_t^0 + \alpha^{1\prime}X_t^1 + e_{t+h} \tag{3}$$

The reason we split X_t into these two components is that it allows the small set of variables X^0 to have a 'large' impact on y_{t+h} through α^0 whereas each element in the high-dimensional vector X^1 has a 'small' impact individually, even though the aggregation of their impacts is non-negligible. This can be related to the above example where the target variable y_{t+h} is inflation, which is predicted strongly by a small set of unemployment-type series, X_t^0, and is lesser affected by the remainder of the big dataset, X_t^1.

Now using the factor model, which we can correspondingly partition into $X_t^0 = \Lambda^0 F_t + u_t^0$ and $X_t^1 = \Lambda^1 F_t + u_t^1$, we can re-write DGP-X in Equation (3) into the alternative formulation:

$$\textbf{(DGP-F)} \qquad\qquad y_{t+h} = \beta' F_t + \alpha^{0\prime} u_t^0 + \alpha^{1\prime} u_t^1 + e_{t+h} \qquad\qquad (4)$$

where $\beta' = \left(\alpha^{0\prime} \Lambda^0 + \alpha^{1\prime} \Lambda^1 \right)$. In the pure factor-augmented case, the methods of Stock and Watson (2002a,b) have an overall regression error term $v_{t+h} = \alpha^{0\prime} u_t^0 + \alpha^{1\prime} u_t^1 + e_{t+h}$ and do not explicitly include u_t^0 in the model. Note that both DGP-X and DGP-F are high dimensional as $N \to \infty$ as they have a number of variables of order N. As noted by Bai and Ng (2009), it is unwise to form predictions using models involving all variables as the mean squared prediction error can be seen to increase in N due to the effect of parameter estimation.

However, it may be preferable to obtain forecasts using small-scale models involving the predictors F_t and/or u_t^0, which are both finite dimensional. We therefore use Equation (2) in making h-step ahead forecasts. Assuming for simplicity that F_t and u_t^0 are known, for the pure factor model case we have $\mathrm{E}\left(y_{T+h}|F_T\right) = \widehat{\beta}' F_T$, whereas for the model including the idiosyncratic components we have $\mathrm{E}\left(y_{T+h}|F_T, u_T^0\right) = \widehat{\beta}' F_T + \widehat{\alpha}^{0\prime} u_T^0$. It can be shown that the difference in mean squared forecast error (MSFE) between the two models is:

$$MSFE\left(y_{T+h}|F_T\right) - MSFE\left(y_{T+h}|F_T, u_T^0\right) =$$
$$\mathrm{E}\left[\alpha^{0\prime} u_T^0 u_T^{0\prime} \alpha^0\right] - \mathrm{E}\left[u_T^{0\prime}\left(\widehat{\alpha}^0 - \alpha^0\right)\left(\widehat{\alpha}^0 - \alpha^0\right)' u_T^0\right] \qquad (5)$$

which can be interpreted as the trade-off between omitting u_t^0 and incurring an unavoidable loss in population, or including u_t^0 in the model and incurring additional parameter estimation error. To give a simple analytical example of this expression, we note as in Bai and Ng (2009) that $T\left(\widehat{\alpha}^0 - \alpha^0\right)\left(\widehat{\alpha}^0 - \alpha^0\right)'$ is $\chi_{m^0}^2$ with expectation m^0 and, if we assume that the idiosyncratic errors are homoskedastic and cross-sectionally uncorrelated with variance σ_u^2 and if the regression error e_{t+h} has variance σ_e^2, then this expression is approximately equal to:

$$\sigma_u^2 \alpha^{0\prime} \alpha^0 - \sigma_e^2 \frac{m^0}{T}$$

Therefore, in cases where α^0 is 'large', for example if $\alpha_i^0 = O\left(1\right)$ for $i = 1, ..., m^0$, then we expect that using the model with idiosyncratic components dominates the pure factor-augmented model in terms of $MSFE$. On the other hand, if α^0 is 'small' or even $\alpha_i^0 = 0$, then there may be no gains in using the additional idiosyncratic components.

Motivated by the empirical and analytical reasons for using the model in Equation (2), the rest of this paper provides results on the optimal selection of the factors and idiosyncratic error terms which enter these types of regression. To the best of our knowledge this is the first paper to look at model selection in this context.

3   Model Selection Criteria and Asymptotic Theory

3.1   Model Selection Criteria

This main difference between our study and previous work is that we additionally have to estimate u_t^0 in Equation (2) as well as F_t^0. Using Principal Components Analysis (PCA) estimation as in Stock and Watson (2002a,b) gives the following feasible analogue to the true model in Equation (2):

$$y_{t+h} = \beta^{0\prime} \widehat{F}_t^0 + \alpha^{0\prime} \widehat{u}_t^{0\prime} + \widehat{\varepsilon}_{t+h} \tag{6}$$

where the regression error $\widehat{\varepsilon}_{t+h}$ is distinguished from the error term ε_{t+h} in the infeasible regression in Equation (2) which contains the latent regressors F_t^0 and u_t^0. The estimated factors $\widehat{F}_t^0$ are the $r^0 \times 1$ subset of the full vector of r estimated factors $\widehat{F}_t^0 \subseteq \widehat{F}_t$, where the $T \times r$ matrix $\widehat{F}$ consists of the r eigenvectors which correspond to the r largest eigenvalues of the $T \times T$ matrix XX', under the identifying normalization $\widehat{F}'\widehat{F}/T = I_r$. This yields the factor loading estimate $\widehat{\Lambda} = X'\widehat{F}/T$. Using both the factor estimates and the factor loadings estimates, along with Equation (1) yields an estimator for each variable i in $\widehat{u}_t^0 \subseteq \widehat{u}_t$ which is equal to:

$$\widehat{u}_{it} = X_{it} - \widehat{\lambda}_i' \widehat{F}_t \tag{7}$$

where $\widehat{\lambda}_i$ corresponds to the ith row of the estimated loadings matrix $\widehat{\Lambda}$. This gives the idiosyncratic part of each variable which is orthogonal to the factors. In papers such as Brownlees et al. (2015), $\widehat{u}_{it}$ represents the 'de-factored' part of the financial asset return X_{it}.

The purpose of our study is to correctly identify the regressors, $\widehat{F}_t^0$ and $\widehat{u}_t^{0\prime}$ by searching over different subsets of $\widehat{F}_t$ and $\widehat{u}_t$. We will refer to a model specification i which uses the variables $\left(\widehat{F}_t^i, \widehat{u}_t^i \right)$ containing r_i estimated factors and m_i estimated idiosyncratic errors. This combination yields the regression:

$$y_{t+h} = \beta^{i\prime} \widehat{F}_t^i + \alpha^{i\prime} \widehat{u}_t^i + \varepsilon_{t+h}^i \tag{8}$$

The class of information criteria we use has a penalty function which depends on both the sample size T and the number of variables N. The difference in this paper with respect to other studies is that the information criterion is a function of estimated idiosyncratic errors as well as factors. For the model i, the criterion to be minimized depends on the number of variables, r_i and m_i, and the estimated sum of squared residuals in model (8):

$$IC \left(\widehat{F}^i, \widehat{u}^i \right) = \ln \left(V \left(\widehat{F}^i, \widehat{u}^i \right) \right) + (r_i + m_i) \, g \left(N, T \right) \tag{9}$$

where:

$$V \left(\widehat{F}^i, \widehat{u}^i \right) = \frac{1}{T} \sum_{t=1}^{T} \left(y_{t+h} - \widehat{\beta}^{i\prime} \widehat{F}_t^i - \widehat{\alpha}^{i\prime} \widehat{u}_t^i \right)^2$$

In order to elaborate on the search procedure, there are a few remarks to make. Firstly, in searching over all subsets of the factors, we note that the total number of r factors in F_t is

typically small, and therefore determining the true subset of r^0 factors F_t^0 in Equation (2) only requires searching over 2^r combinations, which is computationally inexpensive. As mentioned earlier, although r is unknown, it can be estimated in practice using the information criteria of Bai and Ng (2002) in a first step, and then used as the maximum number of factors which we search over in the second-stage forecasting model for y_{t+h}. On the other hand, in selecting variables from u_t, 2^N combinations may not be practical from a computational perspective and, as in standard model selection procedures, we must first reduce the size of the model space over which we search. In other words, we must choose a subset of $m^{max} << N$ variables to search over.

The choice of this candidate set is a decision for the individual researcher. For example, we could be guided by economic theory as in the above example of Stock and Watson (1999) who look at inflation models involving both factors and employment-type series. Therefore we could perform model selection over all r factors, and the subset of m^{max} idiosyncratic components relating to the employment and unemployment series in the dataset. A more general way of generating a candidate subset would be to use a device such as Forward Stepwise (FS) regression and search only over the first m^{max} variables which survive this procedure. Forward Stepwise methods and other sequential methods are surveyed extensively in the chapter of Ng (2013). The principle is to start out with an empty model and add one variable at a time in a way which maximises the fit of the regression. We do not claim to have an optimality result for this type of procedure in this paper, and acknowledge that this has certain shortcomings, as mentioned by Ng (2013). However, since we use this method only as a way to generate a candidate search set we expect it to perform relatively well, and this expectation is confirmed later through Monte Carlo simulation. Future work may address this issue using penalized regression techniques such as LASSO, but this is outside the scope of the current paper.

3.2 Asymptotic Theory

In order to analyse the asymptotic behaviour of the class of information criteria described in Equation (9), we first require some technical assumptions. After a description of these assumptions, we will then show the main result of the paper which is a theorem providing conditions for the selection consistency of these information criteria.

Assumption 1: *The factors satisfy $E\|F_t\|^4 \leq M$ and $\frac{1}{T}\sum_{t=1}^{T}F_t F_t' \underset{p}{\to} \Sigma_F > 0$ as $T \to \infty$.*

Assumption 2: *The factor loadings are either deterministic such that $\|\lambda_i\| \leq M$, or stochastic such that $E\|\lambda_i\|^4 \leq M$, and they satisfy $\frac{1}{N}\Lambda'\Lambda \underset{p}{\to} \Sigma_\Lambda > 0$ as $N \to \infty$.*

Assumption 3: *The idiosyncratic errors are such that: (i) $E(u_{it}) = 0$ and $E|u_{it}|^8 \leq M$ for any i and t, (ii) $E(u_{it}u_{js}) = \sigma_{ij,ts}$ with $|\sigma_{ij,ts}| < \bar{\sigma}_{ij}$ for all t and s, and $|\sigma_{ij,ts}| < \bar{\tau}_{ts}$ for any i and j, and $\frac{1}{N}\sum_{i,j=1}^{N}\bar{\sigma}_{ij} \leq M$, $\frac{1}{T}\sum_{s,t=1}^{T}\bar{\tau}_{ij} \leq M$ and $\frac{1}{NT}\sum_{i,j,t,s=1}\bar{\sigma}_{ij,ts} \leq M$, (iii) $E|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(u_{is}u_{it} - E(u_{is}u_{it}))|^4 \leq M$ for all t and s.*

Assumption 4: *The variables* $\{\lambda_i\}$, $\{F_t\}$ *and* $\{u_{it}\}$ *are three mutually independent groups, although dependence within each group is allowed.*

Assumption 5: *The regression errors satisfy: (i)* $E\left(\varepsilon_{t+h}|\mathcal{F}_t\right) = 0$ *for* $h > 0$ *where* $\mathcal{F}_t = \sigma\left(y_t, F_t, u_t, y_{t-1}, F_{t-1}, u_{t-1}, ...\right)$ *and* $E\left(\varepsilon_{t+h}^4\right) \leq M$, *(ii)* $E|\frac{1}{\sqrt{TN}}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{t+h}\left(u_{is}u_{it} - E\left(u_{is}u_{it}\right)\right)|^2 \leq M$ *for each* s *and for* $h > 0$, *(iii)* $E\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_i u_{it}\varepsilon_{t+h}\|^2 \leq M$ *with* $E\left(\lambda_i u_{it}\varepsilon_{t+h}\right) = 0$ *for any* i *and* t, *and (iv)* $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Z_t\varepsilon_{t+h} \underset{d}{\rightarrow} N\left(0,\Omega\right)$ *where* $Z_t = \left[F_t^0, u_t^0\right]$.

Assumptions 1-3 are standard in the literature of factor-augmented forecasting models for ensuring the existence of r factors, and allowing for heteroskedasticity and limited dependence in the idiosyncratic errors. These assumptions coincide with those of Bai and Ng (2006), with the exception of their Assumption C4 which they require to derive the asymptotic distribution of the factors, which is not required in this paper. Assumption 4 requires mutual independence of λ_i, F_t and u_{it} as in Bai and Ng (2006), though can be readily replaced with weaker assumptions along the lines of Gonçalves and Perron (2014). Finally, Assumption 5 follows Cheng and Hansen (2015) in placing standard moment restriction on the factor-augmented model, in this case modified to include the idiosyncratic components.

Selection consistency occurs when the probability limits of the generated regressors span the same space as the true factors and idiosyncratic errors F_t^0 and u_t^0. For the factors, it is well known that the Principal Components estimates converge to a particular rotation of the true factors $H^0 F_t^0$, as shown by Bai and Ng (2002). On the other hand, the estimates of the idiosyncratic errors $\widehat{u}_t^i$ are consistent for the true u_t^i without rotation.

The following theorem shows the conditions on $g\left(N,T\right)$ required for consistency of selection:

Theorem 1. *Let Assumptions 1-5 hold and let the factors and factor loadings be estimated by Principal Components. For two models* i *and* j, *if model* i *corresponds to the true model such that the probability limit of* $\left(\widehat{F}^i, \widehat{u}^i\right)$ *is* $\left(H^0 F_t^0, u_t^0\right)$ *for all* t, *and for model* j *one or both of* $\widehat{F}_t^j$ *and* $\widehat{u}_t^j$ *has different probability limit, then:*

$$\lim_{N,T\rightarrow\infty} \Pr\left(IC\left(\widehat{F}^j, \widehat{u}^j\right) < IC\left(\widehat{F}^i, \widehat{u}^i\right)\right) = 0$$

as long as (i) $g\left(N,T\right) \rightarrow 0$ *and (ii)* $\min\left\{\sqrt{T}, N\right\} g\left(N,T\right) \rightarrow \infty$ *as* $N,T \rightarrow \infty$.

The proof of this result and other required results can be found in the separate Online Appendix. The key difference of this result relative to previous information criteria is that Condition (ii) requires that $\min\left\{\sqrt{T}, N\right\} g\left(N,T\right) \rightarrow \infty$ rather than $\min\left\{T, N\right\} g\left(N,T\right) \rightarrow \infty$, which was the condition required in pure factor model studies such as Bai and Ng (2009) and Groen and Kapetanios (2013).[5] The cause of this difference is a result which can be seen in Lemma 1 in

---

[5]Condition (i) is standard for information criteria in ensuring that inflation in $V\left(.\right)$ due to incorrect model specification is always larger than the penalty $g\left(N,T\right)$. We therefore do not discuss this condition further.

the Online Appendix. This Lemma shows that the product of the estimation error in the idiosyncratic component, $(\widehat{u}_{it} - u_{it})$, with the factors vanishes at the rate $\min\left\{\sqrt{T}, N\right\}$, rather than the $\min\{T, N\}$ rate obtained for the factor estimation error in previous studies.

The main implication of this result is that existing information criteria used in previous studies do not meet the requirements in Condition (ii) for consistent model selection, and should therefore not be used in specifying models which take this form. In particular, even model selection criteria which are modified to allow for factor estimation error are inconsistent as they only take the $\min\{T, N\}$ consistency rate of the factors into account. These criteria are subsequently inconsistent in all cases where $\sqrt{T}$ is of smaller order than N; a scenario which can happen quite readily in macroeconomic datasets where the panel dimensions are roughly the same. As will be shown in the next section, information criteria such as those in Groen and Kapetanios (2013), do not satisfy Condition (ii) required by Theorem 1 which leads to inconsistent model selection. It is also useful to note that, while this result holds for the consistency rates resulting from the Principal Components factor estimates, in the same way as Groen and Kapetanios (2013) the second condition can be modified to reflect the consistency rates of other factor model estimators such as the maximum likelihood method of Doz et al. (2011, 2012). The relative performance of our criteria to other criteria will be similar after applying any modifications to these consistency rates.

In specifying Equation (2) as a function of both F_t^0 and u_t^0, it was implicitly assumed that both the factors and idiosyncratic components are relevant in the DGP for y_{t+h}. While the result of Theorem 1 still holds when either $\alpha^0 = 0$ or $\beta^0 = 0$, there is an important further remark to make about the latter case where $\beta^0 = 0$ and the factors are irrelevant in the true DGP for y_{t+h}:

Remark 1. *In the case where $\beta^0 = 0$ and the factors are irrelevant in the true DGP for y_{t+h}, then the results of Theorem 1 still hold, but it is possible to replace Condition (ii) with the following modified condition: (ii') $\min\{T, N\}\, g(N, T) \to \infty$ as $N, T \to \infty$.*

This remark shows that, in the eventuality that the factors do not appear in the true DGP for y_{t+h}, then we have the same conditions required on the penalty function as in the pure factor augmented case.[6] This would suggest that if it is known *a priori* that $\beta^0 = 0$, then we may use the model selection methods of Bai and Ng (2009) and Groen and Kapetanios (2013), even though we are using the estimated idiosyncratic components in the model. This could, in theory, lead to some finite sample improvements. However, given that it is generally not known *a priori* whether $\beta^0 = 0$ in practice, and since Condition (ii) in Theorem 1 is sufficient for Condition (ii') in Remark 1 in that $\min\left\{\sqrt{T}, N\right\} g(N, T) \to \infty$ implies $\min\{T, N\}\, g(N, T) \to \infty$, it is more reliable to use information criteria which satisfy Condition (ii) rather than Condition (ii').

In the next section we propose a new set of information criteria which satisfy Conditions (i) and (ii) of Theorem 1 and compare these to existing information criteria which we show to be inconsistent.

---

[6]We thank an anonymous referee for this helpful comment.

4   Information Criteria

In satisfying the conditions imposed on the penalty function $g(N,T)$ by Theorem 1, we propose several different information criteria based both on a Mallows-type form and a Hannan-Quinn form. The penalty functions we propose make use of the fact that $\max\left\{\frac{1}{\sqrt{T}},\frac{1}{N}\right\}\approx\frac{\sqrt{T}+N}{\sqrt{T}N}$. Since we ensure that all of the criteria satisfy both Conditions (i) and (ii) of Theorem 1, they are all consistent and therefore equivalent asymptotically. However, their performance may differ in finite samples. These finite sample properties will be assessed later through Monte Carlo simulations.

The new information criteria are as follows:

$$IC_1\left(\widehat{F}^i,\widehat{u}^i\right) = \ln\left(V\left(\widehat{F}^i,\widehat{u}^i\right)\right) + (r_i+m_i)\ln\left(\frac{\sqrt{T}N}{\sqrt{T}+N}\right)\left(\frac{\sqrt{T}+N}{\sqrt{T}N}\right)$$

$$IC_2\left(\widehat{F}^i,\widehat{u}^i\right) = \ln\left(V\left(\widehat{F}^i,\widehat{u}^i\right)\right) + (r_i+m_i)\ln\left(\min\left\{\sqrt{T},N\right\}\right)\left(\frac{\sqrt{T}+N}{\sqrt{T}N}\right)$$

$$IC_3\left(\widehat{F}^i,\widehat{u}^i\right) = \ln\left(V\left(\widehat{F}^i,\widehat{u}^i\right)\right) + (r_i+m_i)\frac{\ln\left(\min\left\{\sqrt{T},N\right\}\right)}{\min\left\{\sqrt{T},N\right\}}$$

$$HQ_1\left(\widehat{F}^i,\widehat{u}^i\right) = \ln\left(V\left(\widehat{F}^i,\widehat{u}^i\right)\right) + 2(r_i+m_i)\ln\ln\left(\frac{\sqrt{T}N}{\sqrt{T}+N}\right)\left(\frac{\sqrt{T}+N}{\sqrt{T}N}\right)$$

$$HQ_2\left(\widehat{F}^i,\widehat{u}^i\right) = \ln\left(V\left(\widehat{F}^i,\widehat{u}^i\right)\right) + 2(r_i+m_i)\ln\ln\left(\min\left\{\sqrt{T},N\right\}\right)\left(\frac{\sqrt{T}+N}{\sqrt{T}N}\right)$$

$$HQ_3\left(\widehat{F}^i,\widehat{u}^i\right) = \ln\left(V\left(\widehat{F}^i,\widehat{u}^i\right)\right) + 2(r_i+m_i)\frac{\ln\ln\left(\min\left\{\sqrt{T},N\right\}\right)}{\min\left\{\sqrt{T},N\right\}}$$

These criteria clearly satisfy both Conditions (i) and (ii) of Theorem 1. In comparing these criteria to previous literature, we firstly note that the penalty functions used in IC_1, IC_2 and IC_3 criteria are similar in nature to those in Bai and Ng (2002). However, as the information criteria of Bai and Ng (2002) were for selecting the number of factors in the first-stage factor model and not in the second-stage forecasting regression, we will not make a direct comparison of our results to theirs. The Hannan-Quinn criteria HQ_1, HQ_2 and HQ_3 are similar in spirit to those in Groen and Kapetanios (2013). As such, we will now discuss the differences of our criteria to those, with particular reference to the conditions for selection consistency shown in Theorem 1.

The $BICM$ and $HQICM$ criteria of Groen and Kapetanios (2013) were proposed for the case of pure factor-augmented model selection, with no estimated idiosyncratic components.[7] Their information criteria, applied to the forecasting model in Equation (2), are as follows:

$$BICM\left(\widehat{F}^i,\widehat{u}^i\right) = \ln\left(V\left(\widehat{F}^i,\widehat{u}^i\right)\right) + (r_i+m_i)\ln(T)\left(\frac{T+N}{TN}\right)$$

---

[7]In their paper they scale the criteria by T whereas in this paper we follow more closely the specifications of Bai and Ng (2002).

$$HQICM\left(\widehat{F}^i, \widehat{u}^i\right) = \ln\left(V\left(\widehat{F}^i, \widehat{u}^i\right)\right) + 2\left(r_i + m_i\right)\ln\ln\left(T\right)\left(\frac{T+N}{TN}\right)$$

While both of these criteria have $g\left(N,T\right) \to 0$ and therefore pass our Condition (i), they both fail Condition (ii) in cases where $\sqrt{T}$ is of smaller order than N, which includes cases such as that in which $N = T$ as $N, T \to \infty$. This means that in finite samples, panels where N is of similar or larger order than $\sqrt{T}$, we expect our proposed criteria to provide significant improvements over both of the methods proposed by Groen and Kapetanios (2013). On the other hand, in the case of Remark 1, where the factors are not relevant for y_{t+h}, Condition (ii') coincides with the condition required for consistency in Groen and Kapetanios (2013), and the $BICM$ and $HQICM$ criteria are consistent.

It is also useful to compare our criteria to the standard AIC and BIC criteria which are commonly used in time series applications:

$$AIC\left(\widehat{F}^i, \widehat{u}^i\right) = \ln\left(V\left(\widehat{F}^i, \widehat{u}^i\right)\right) + \left(r_i + m_i\right)\frac{2}{T}$$

$$BIC\left(\widehat{F}^i, \widehat{u}^i\right) = \ln\left(V\left(\widehat{F}^i, \widehat{u}^i\right)\right) + \left(r_i + m_i\right)\frac{\ln\left(T\right)}{T}$$

Both the AIC and the BIC also fulfil Condition (i) as $g\left(N,T\right) \to 0$ but they both fail Condition (ii) for all configurations of N and T and are therefore inconsistent.[8] This result is somewhat alarming as the vast majority of empirical forecasting studies tend to use the BIC for model selection. In our simulations, we therefore expect there to be overparameterization in all Monte Carlo specifications of N and T for these criteria. This result is unusual relative to the pure factor-augmented approach of Groen and Kapetanios (2013) as in that set-up the BIC is still consistent in cases where $N << T$. In our case, the presence of the estimated idiosyncratic components drives inconsistency in the BIC, and we must use the modified the penalty functions proposed above.

5 Monte Carlo

5.1 Data Generating Process

In this section we specify the Monte Carlo Data Generating Processes (DGPs) which we will use to obtain results on the finite sample performance of the information criteria proposed above. We follow the structure proposed in Section 2, where we obtain the forecasting model in Equation (2) first by specifying a DGP for y_{t+h} as a function of X_t, as in DGP-X in Equation (3), and then by substituting in the factor model in Equation (1) for X_t. This allows us to mimic the situation described above, where the true DGP for the variable of interest is related with different strength to the observable variables, X_t, and with the factors being a purely statistical device for data

---

[8]In a related paper, Carrasco and Rossi (2016) note that their cross-validation and Mallows-type criteria are AIC-type criteria and would also be inconsistent in this setting.

reduction. This is the same approach as taken by Groen and Kapetanios (2013). Specifically:

$$y_t = \alpha^{0\prime} X_t^0 + \alpha^{1\prime} X_t^1 + \sqrt{\theta}\varepsilon_t \tag{10}$$

$$X_t = \frac{1}{\sqrt{r}}\Lambda F_t + u_t \tag{11}$$

where X_t^0 and X_t^1 are sub-vectors of X_t of dimension $m^0 \times 1$ and $\left(N - m^0\right) \times 1$ respectively. Splitting X_t in this way enables us to specify 'large' values for the coefficients α^0 and 'small' values for the coefficients α^1 so that we arrive at the forecasting model of Equation (2):

$$y_t = \underbrace{\frac{1}{\sqrt{r}}\left(\alpha^{0\prime}\Lambda^0 + \alpha^{1\prime}\Lambda^1\right)F_t + \alpha^{0\prime}u_t^0}_{\text{Explained}} + \underbrace{\left(\alpha^{1\prime}u_t^1 + \sqrt{\theta}\varepsilon_t\right)}_{\text{Unexplained}}$$

We let the 'large' and 'small' coefficients be $\alpha^0 = \mathbf{1}_{m^0 \times 1}$ and $\alpha^1 = \mathbf{1}_{(N-m^0)\times 1}/\sqrt{N - m^0}$.[9] The aim of the model selection is therefore to include the first m^0 idiosyncratic components in the model as they have a large impact on y_t, with the remaining negligible error terms u_t^1 entering the unexplained part of the regression. This is unlike in Groen and Kapetanios (2013) where the unexplained part of the pure factor-augmented model contains all N idiosyncratic errors. No *a priori* knowledge is assumed of the split between u_t^0 and u_t^1; as explained in Section 3, a candidate search set to uncover u_t^0 is first specified using Forward Stepwise regression.

The regression errors ε_t are drawn such that $\varepsilon_t \sim i.i.d. N\left(0, 1\right)$. The idiosyncratic errors are also drawn from a normal distribution, but the variance differs between the first m^0 variables and the remaining variables, with $u_{it} \sim i.i.d. N\left(0, K\right)$ for $i \leq m^0$ and $u_{it} \sim i.i.d. N\left(0, 1\right)$ for $i > m^0$. Finally, the factor loadings are drawn with a non-zero mean such that $\Lambda \sim i.i.d. N\left(1, 1\right)$ and the factors are $F_t \sim i.i.d. N\left(0, 1\right)$. The rescaling by $1/\sqrt{r}$ in Equation (11) ensures that the variance of X_{it} is K for $i \leq m^0$ and 1 for $i > m^0$. We then choose the parameters K and θ to fix the signal to noise ratio in Equation (10), and also to equate the signal to noise ratio between F_t and u_t^0.

Given the distributions of Λ, F, u and ε, the overall regression R^2 is:

$$
\begin{aligned}
R^2 &= 1 - \frac{\alpha^{1\prime}\alpha^1 + \theta}{\alpha^{0\prime}\alpha^0 + \alpha^{1\prime}\alpha^1 + K\alpha^{0\prime}\alpha^0 + \alpha^{1\prime}\alpha^1 + \theta} \\
&= 1 - \frac{1 + \theta}{m^0 + 1 + Km^0 + 1 + \theta}
\end{aligned}
$$

We therefore set $K = \left(m^0 + 1\right)/m^0$ to equate the variation of y_t explained by F_t and u_t^0 and we set θ so that $R^2 = 1/2$ which requires that $\theta = 1 + 2m^0$.

For the sample sizes, we consider a fixed set of values for the time series dimension corresponding to $T = 50, 100, 200, 400$. For the cross section dimension we consider different asymptotic rules corresponding to $N = c_1 T$ and $N = c_2\sqrt{T}$. These two particular asymptotic set-ups are chosen

---

[9]Therefore when $N = 50$ and $m^0 = 1$ the coefficient on the first variable is 7 times the size of the remaining coefficients.

because, in the case where $N = c_1 T$, as mentioned in Section 4, we expect consistent model selection only for the new information criteria proposed in this paper, although in finite samples the pure factor augmented model criteria of Groen and Kapetanios (2013) may perform reasonably when N is small. When $N = c_2\sqrt{T}$, we expect inconsistency of the AIC and the BIC, though the criteria of Groen and Kapetanios (2013) should be consistent along with those suggested in this paper.

We therefore select 2 different levels of c_1 and c_2 which fix the level of N equal to 20 and 50 when $T = 50$. We call these the "Small N" and "Large N" scenarios:[10]

Table 1: Scenarios for Sample Sizes of T and N

	"Small N"		"Large N"	
Scenario:	1	2	3	4
T	$N = 0.4T$	$N = 2.83\sqrt{T}$	$N = T$	$N = 7.07\sqrt{T}$
50	20	20	50	50
100	40	28	100	71
200	80	40	200	100
400	160	57	400	141

Notes: The coefficients multiplying T and $\sqrt{T}$ in the expressions for N are set so that $N = 20, 50$ when $T = 20$ in the "Small" and "Large" scenarios. Note that $2.83 \approx 20/\sqrt{50}$ and $7.07 \approx 50/\sqrt{50}$.

We are interested in two aspects of the results: that the correct number of variables are selected, and that the identity of their selection is correct, both for factors and idiosyncratic errors. From the results in Theorem 1 we expect that the AIC and BIC criteria overestimate r^0 and m^0 for all combinations of T and N and that the $MBIC$ and $HQICM$ of Groen and Kapetanios (2013) overestimate in cases where $T << N$.

In order to assess the number of selected variables, we will use the average number of selected idiosyncratic components and factors, $\widehat{m}^0$ and $\widehat{r}$, over $B = 1000$ Monte Carlo replications. To assess variable selection, we use a mean squared deviation (MSD) statistic. If we denote $\widehat{S}^F$ and $\widehat{S}^u$ as the $r \times 1$ and $N \times 1$ binary selection vectors of 1's and 0's which correspond to the minimization of a given criterion, and S_0^F and S_0^u are the true inclusion vectors according to the data generating process in Equation (10) then the MSD statistics for F and u are:

$$MSD^F = \frac{1}{B} \sum_{b=1}^{B} \left(\widehat{S}_b^F - S_0^F\right)' \left(\widehat{S}_b^F - S_0^F\right) \tag{12}$$

$$MSD^u = \frac{1}{B} \sum_{b=1}^{B} \left(\widehat{S}_b^u - S_0^u\right)' \left(\widehat{S}_b^u - S_0^u\right) \tag{13}$$

where b indexes the Monte Carlo replication. Values of MSD equal to zero therefore represent

---

[10]These combinations of N and T are typical of many of the types of dataset commonly used in applied forecasting studies. Further results for other combinations of sample size are available upon request.

perfect model selection, whereas large values of MSD represent poor model selection.

In the following section we document the results for the scenario where $r = 1$ and we let $m^0 = 1, 2$. In practice for the simulations, the BIC_3 criterion of Bai and Ng (2002) is used to estimate the number of factors present in the first stage and to obtain the estimated idiosyncratic components. In the forecasting equation, however, we search over a maximum possible set of $r^{max} = 5$ factors and $m^{max} = 5$ idiosyncratic errors in order to allow the possibility of over-selection of redundant factors. The largest model we consider in the search procedure therefore contains 10 variables.

5.2 Results

Figures 1 to 4 show the number of chosen idiosyncratic errors, $\widehat{m}^0$, and factors, $\widehat{r}$, averaged across the 1,000 Monte Carlo replications for the 2 specifications of $r = 1$ and $m^0 = 1, 2$. The results for the Mean Squared Deviation statistics MSD^u and MSD^F for the same parameter configurations, along with the results for further parameter values, can be found in the separate Online Appendix.

[Display Figures 1 to 4 here]

The results clearly demonstrate the selection consistency of the newly proposed information criteria IC_1, IC_2, IC_3, HQ_1, HQ_2 and HQ_3. There are a few key features of the results to highlight in particular. Firstly, consistency occurs over all different combinations of N and T considered in Scenarios 1 through 4. This is unlike the cases of the AIC and BIC which display inconsistent model selection in all configurations, and the $BICM$ and $HQICM$ criteria of Groen and Kapetanios (2013), where overfitting occurs in the cases where N is large relative to T. Secondly, while the new criteria behave the same way asymptotically, it appears from these results that the Hannan-Quinn type criteria perform better than IC_1, IC_2 and IC_3, with HQ_3 having the best finite sample performance with MSD figures closest to 0, as can be seen in the Online Appendix. Thirdly, all six new criteria have particularly strong selective power with respect to the factors, even for very low T and N. This is in contrast to the $BICM$ and $HQICM$ criteria when N is large. Finally, we see that as the number of idiosyncratic terms m^0 increases from 1 to 2, selection worsens and the criteria become more conservative. This is another fairly standard result, although the HQ_3 criteria still appears to perform well when the panel dimensions are reasonable.

Turning to the other information criteria, as mentioned above, we can see that both the AIC and BIC in Figures 1 to 4 overfit the model for all values of N and T, both for the selection of factors and variables. In many cases the number of selected factors and variables is close to the maximum number considered in the search procedure which is 5. This result is unsurprising for the AIC which is known to be inconsistent in all standard set-ups. The result for the BIC, as mentioned before, is somewhat unusual as in the pure-factor augmented set-up of Bai and Ng (2009) and Groen and Kapetanios (2013), where the search takes place only over factor estimates, the BIC is consistent for $N << T$. In our set-up the BIC is inconsistent in *all* cases; a result which is due to the search over additional estimates of the idiosyncratic error terms.

The results for the $BICM$ and $HQICM$ criteria of Groen and Kapetanios (2013) illustrate the property shown in the previous section, that these criteria for pure factor-augmented models are only consistent for cases where N is small relative to $\sqrt{T}$. This can be seen from Scenarios 1 and 2 in Figures 1 to 4 which correspond to the "Small N" scenario, in which the $BICM$ and $HQICM$ perform reasonably. However, in the cases of "Large N" the inconsistency of these criteria under Scenarios 3 is clear, indicating that care must be exercised if applying these criteria when estimated idiosyncratic components are used in the model. Even in Scenario 4, where we expect $BICM$ and $HQICM$ to give consistent model selection, their finite sample performance tends to be worse than the new criteria IC_1 to HQ_3.

Finally, we also ran a set of results where $m^0 = 0$ in order to mimic the situation where we incorrectly search over idiosyncratic components when the true DGP is, in fact, the pure factor-augmented model.[11] These results, displayed in the Online Appendix, show that our criteria remain successful at estimating $m^0 = 0$. Although performance is slightly worse than the cases where $m^0 > 0$, particularly for Scenario 2, our criteria perform significantly better than the other criteria we use for comparison.

6   Empirical Application: Long-Horizon Exchange Rate Modelling

In this section we provide an empirical application of these model selection procedures in a very challenging predictive environment: long-horizon exchange rate modelling. We extend a recent approach of Engel et al. (2015) who suggest to predict exchange rate growth using the idiosyncratic component from a factor model of countries' exchange rates. This model provides a natural application for evaluating our new consistent model selection criteria.

6.1   Introduction and Background

Since the seminal work of Meese and Rogoff (1983b,a), there has been growing interest in the predictive ability of log exchange rates over long forecast horizons. There have been many subsequent empirical studies looking to relate the growth of exchange rates at different horizons to macroeconomic 'fundamentals'. These approaches are comprehensively surveyed in Rossi (2013). The standard empirical approach in the literature tends to follow that of Mark (1995) and Kilian (1999) in specifying regression models such as:

$$s_{i,t+h} - s_{it} = \mu + \beta \left(s_{it} - z_{it} \right) + \varepsilon_{i,t+h} \tag{14}$$

where s_{it} is the logarithm of the exchange rate of country i, usually relative to the U.S. dollar (\$), and z_{it} is some measure of central tendency for s_{it}, or 'fundamental'. Various different fundamentals have been considered for z_{it} in the literature, corresponding to different macroeconomic models of

---

[11]We thank an anonymous referee for the suggestion to explore this further.

exchange rates. The most commonly used fundamentals are variables such as trade balances, inflation, income and money supply.

Recently, the paper of Engel et al. (2015) moved away from the standard set of fundamentals used in the literature. They suggest to use a factor model for a panel of global exchange rates relative to the U.S. dollar (\$), denoted $s_t = \Lambda F_t + u_t$, and to use the common component for each country $i = 1, ..., N$ in place of the fundamental. In other words, they suggest to replace z_{it} with $\lambda_i' F_t$, where λ_i is the factor loading vector for country i, corresponding to the ith row of Λ. Having estimated Λ and F_t, for a given country i they use the estimated idiosyncratic component $\widehat{u}_{it} = s_{it} - \widehat{\lambda}_i' \widehat{F}_t$ and substitute this directly into the forecasting Equation (14) in place of $(s_{it} - z_{it})$. Therefore their model has exactly the same form of regression model which is the focus of this paper. Our proposed information criteria may therefore be used to select between alternative specifications.

In this application we choose to go one step further than Engel et al. (2015) and use the full set of (cross-country) idiosyncratic errors to allow for potential exchange rate spillover effects from other economies, all relative to the common factor in the panel of exchange rates. Empirical studies into exchange rate spillover in both mean and variance has been considered by papers such as Hong (2001) and others. These spillover effects, if present, will be selected using the information criteria proposed in this paper. We therefore combine the literatures of long-horizon exchange rate forecasting, factor models and exchange rate spillovers.

Specifically, in the next section we will look at in-sample model selection in a baseline regression which takes the following form:

$$s_{i,t+h} - s_{it} = \mu + \alpha_i \widehat{u}_{it} + \sum_{j \neq i} \alpha_j \widehat{u}_{jt} + \varepsilon_{i,t+h} \tag{15}$$

where $\widehat{u}_{it}$ is the domestic idiosyncratic error as in Engel et al. (2015), whereas $\widehat{u}_{jt}$, for $j \neq i$, is the exchange rate spillover effect new to this paper. Therefore, when $\alpha_j = 0$, Equation (15) corresponds exactly to the specification of Engel et al. (2015). We will also perform some robustness checks involving interactions of the idiosyncratic components with recession dummy variables.

In the last section, we will perform an analysis of this forecasting model in a pseudo out-of-sample forecasting context. The evidence on out-of-sample predictive ability of long-horizon exchange rate models is very mixed in the literature. The assertion of Meese and Rogoff (1983b,a) was that no model could outperform the naïve no-change benchmark model. Since their work, there has been a great many papers attempting to overturn this result, for example Mark (1995). The paper of McCracken and Sapp (2005) also suggested that, using a range of different models and a range of predictive ability test statistics, there is some positive evidence of beating the no-change forecasting model. In addition to model (15), we will explore different model specifications using a larger model which also includes a macroeconomic fundamental:

$$s_{i,t+h} - s_{it} = \mu + \alpha_i \widehat{u}_{it} + \sum_{j \neq i} \alpha_j \widehat{u}_{jt} + \beta (s_{it} - z_{it}) + \varepsilon_{i,t+h} \tag{16}$$

For the macroeconomic fundamental, z_{it}, we will use the PPP model since this outperforms other models such as Taylor-rules in Engel et al. (2015). In light of previous empirical evidence, we are not too hopeful of very positive results from an out-of-sample perspective. However, it will be useful to see how parsimonious our model selection procedures are in a pseudo out-of-sample setting.

6.2 Data and Estimation

To form the factor dataset, we choose the same countries as used in Engel et al. (2015). We use monthly, rather than quarterly, data for the average closing exchange rate of 18 OECD countries plus the Eurozone relative to the U.S. Dollar over the time period August 1988 to May 2015. The data we use to construct the price level differential for the PPP model is the Consumer Price Index (CPI), which is available at a monthly level for all countries except for Australia. All data is extracted from the Haver Analytics databases USECON, G10 and EMERGE.[12] We split the sample at the end of 1998 and will perform the analysis for the pre- and post-Euro sub samples. The countries are listed in Table 2.

Table 2: List of countries in the dataset, split by pre-Euro and post-Euro sub-samples with cut-off date December 1998.

Period	Countries
Pre-Euro	Australia (AUS), Austria (AUT), Belgium (BEL), Canada (CAN), Denmark (DNK), Finland (FIN), France (FRA), Germany (DEU), Italy (ITA), Japan (JPN), Korea (KOR), Netherlands (NLD), Norway (NOR), Spain (ESP), Sweden (SWE), Switzerland (CHE), United Kingdom (GBR)
Post-Euro	Australia (AUS), Canada (CAN), Denmark (DNK), Europe (EUR), Japan (JPN), Korea (KOR), Norway (NOR), Sweden (SWE), Switzerland (CHE), United Kingdom (GBR)

This means that for the pre-Euro sub-sample the dataset is of size $T = 197$, $N = 17$ and for the post-Euro sub-sample we have $T = 137$, $N = 10$. We will look at the monthly forecast horizons $h = 1, 3, 6, 9, 12, 18, 24$, which is similar to other studies such as McCracken and Sapp (2005) and Engel et al. (2015) who use quarterly data and a horizon of 2 or 3 years.

In the first-stage factor model, we estimate the factors and idiosyncratic components by Principal Components, and let the number factors be estimated by the BIC_3 criterion of Bai and Ng (2002). This is in contrast to Engel et al. (2015) who fix an arbitrary number of factors such as $r = 2$, and compare the results to other factor specifications.[13] In this paper we prefer to use a consistent criterion to select the number of factors, as under-estimation of r can cause the idiosyncratic errors to be incorrectly correlated, which would invalidate the results of our model selection procedures.

---

[12]Data accessed 18th June 2015. We thank Now-Casting Economics Ltd. for access to this data source.

[13]In an earlier version of the paper, we presented similar results which held the number of factors fixed throughout the study. These results are available upon request.

In the second stage model, Equation (15), we will perform model selection using our proposed HQ_3 criterion up to a maximum possible number of $m^{max} = 5$ idiosyncratic error spillover components. An important remark to make is that the model approach of Engel et al. (2015) makes use of only the idiosyncratic components, which appears to impose that the factors themselves do not appear in the DGP for the log exchange rate growth; a situation mentioned in Remark 1. As mentioned in Section 3, we prefer to use the HQ_3 criterion from this paper rather than one of the criteria of Groen and Kapetanios (2013) as it provides consistent model selection regardless of the unknown DGP or sample size. However, we will also run checks to see whether our results are different to the results when using other selection criteria.

6.3   In-Sample Model Selection Results

We first present the model selection results using all available data in both the pre- and post-Euro sub-samples. Our interest is whether the new model selection criteria select any spillover idiosyncratic components $\widehat{u}_{jt}$ in model (15) over and above the domestic idiosyncratic component $\widehat{u}_{it}$. There are several reasons why we may expect very few additional variables to be chosen. Firstly, the sample size is small in the cross section dimension, N, for which we know our selection criteria can be slightly conservative. Secondly, we know from previous evidence that additional variables tend to have weak predictive power over and above the "no-change" benchmark, which may mean that few additional variables will be selected.

Tables 3 and 4 present the model selection results for the spillover effects in the pre- and post-Euro sub-samples respectively. We present only the results for the $h = 1$ and $h = 12$ horizons for brevity. There are several key features to highlight from these results. The most surprising result is that, particularly in the pre-Euro era which includes the major European economies, the consistent HQ_3 criterion displays a non-trivial selection of spillover effects for the $h = 12$ horizon. In most cases there are between 1 and 3 spillover effects chosen. This is contrary to the intuition that the model selection criterion would be over-conservative, both because of the small sample size, and because of the harsh predictive environment. In many cases, the selected spillover effects also have some reasonable interpretation, even though there is no theoretical model of spillovers in place. For example, we see linkages between the major European economies such as Austria and Switzerland, and the Scandinavian countries. However, due to the small panel we consider, the results should also be treated with caution. For example, the countries with the most number of idiosyncratic components selected are Canada and Korea which are somewhat isolated within our sample and, as such, the model selection criterion appears to deliver spillovers which may not be deemed sensible. This gives motivation for a more extensive study involving a larger panel of global economies.

In the post-Euro era, however, much fewer spillover effects are chosen at the $h = 12$ horizon by the consistent HQ_3 criterion. This may be in some part due to the small number of countries in this sub-sample, $N = 10$. This is a number smaller than we considered in the Monte Carlo simulations, and is likely to have quite conservative model selection. The other main features of the results are that, at the short $h = 1$, there are no spillover effects chosen by the HQ_3 criterion. This is in line

Table 3: In-Sample Model Selection Results: 1988-1998 pre-Euro era.

	Selection Criterion: HQ_3	
	$h = 1$	$h = 12$
	Countries ($\widehat{m}$)	Countries ($\widehat{m}$)
AUS	- (0)	CHE (1)
AUT	- (0)	- (0)
BEL	- (0)	DNK (1)
CAN	- (0)	FIN, ESP, BEL (3)
DNK	- (0)	BEL (1)
FIN	- (0)	- (0)
FRA	- (0)	NOR (1)
DEU	- (0)	- (0)
ITA	- (0)	FIN, CAN (2)
JPN	- (0)	ITA, BEL (2)
KOR	- (0)	DNK, ESP, CHE (3)
NLD	- (0)	- (0)
NOR	- (0)	DNK (1)
ESP	- (0)	FIN (1)
SWE	- (0)	FIN, ESP (2)
CHE	- (0)	- (0)
GBR	- (0)	FIN (1)
	Selection Criterion: BIC	
	$h = 1$	$h = 12$
	Countries ($\widehat{m}$)	Countries ($\widehat{m}$)
AUS	CHE (1)	CHE, ITA, FIN, BEL, ESP (5)
AUT	CAN, BEL, DNK (3)	NOR, DNK, BEL, FIN (4)
BEL	DNK, JPN (2)	DNK, NOR (2)
CAN	SWE, GBR, FRA (3)	FIN, ITA, ESP, BEL (4)
DNK	JPN, BEL (2)	NOR, FIN, CHE (3)
FIN	ITA (1)	SWE, ITA (2)
FRA	DNK, JPN, BEL (3)	NOR, DNK, FIN, CHE (4)
DEU	BEL, CAN, DNK (3)	NOR, DNK, FIN, BEL (4)
ITA	NOR, FIN (2)	FIN, CAN (2)
JPN	ITA (1)	ITA, BEL (2)
KOR	ITA, FIN (2)	DNK, ESP, CHE, SWE, JPN (5)
NLD	CAN, BEL, DNK (3)	NOR, DNK, BEL (3)
NOR	CAN (1)	DNK, FIN, CHE, ITA (4)
ESP	FIN, BEL, DNK, JPN (4)	FIN, DEU, ITA, GBR (4)
SWE	CHE (1)	FIN, ITA, ESP (3)
CHE	BEL, FIN, ITA (3)	NOR, DNK, FIN (3)
GBR	AUS, SWE (2)	FIN, FRA (2)

Notes: For each country, the column "Countries ($\widehat{m}$)" displays the identity of selected spillover countries, and the number of these selected countries in parentheses. Model selection is performed using the HQ_3 criterion (upper panel) and BIC criterion (lower panel). The results are displayed only for the horizons $h = 1$ and $h = 12$, with other results available upon request.

Table 4: In-Sample Model Selection Results: 1999-2015 post-Euro era.

	Selection Criterion: HQ_3	
	$h = 1$	$h = 12$
	Countries ($\widehat{m}$)	Countries ($\widehat{m}$)
AUS	- (0)	SWE (1)
CAN	- (0)	SWE (1)
DNK	- (0)	SWE (1)
EUR	- (0)	SWE (1)
JPN	- (0)	- (0)
KOR	- (0)	- (0)
NOR	- (0)	SWE (1)
SWE	- (0)	- (0)
CHE	- (0)	CAN (1)
GBR	- (0)	- (0)
	Selection Criterion: BIC	
	$h = 1$	$h = 12$
	Countries ($\widehat{m}$)	Countries ($\widehat{m}$)
AUS	- (0)	SWE, CHE (2)
CAN	SWE (1)	SWE (1)
DNK	CAN (1)	NOR, CAN, CHE, EUR (4)
EUR	CAN (1)	SWE, NOR, CAN (3)
JPN	CAN (1)	SWE, AUS, GBR (3)
KOR	- (0)	CHE, NOR, SWE, AUS (4)
NOR	- (0)	SWE, KOR (2)
SWE	NOR (1)	NOR, AUS, CHE (3)
CHE	CAN (1)	JPN, SWE, EUR (3)
GBR	- (0)	SWE, CHE (2)

Notes: For each country, the column "Countries ($\widehat{m}$)" displays the identity of selected spillover countries, and the number of these selected countries in parentheses. Model selection is performed using the HQ_3 criterion (upper panel) and BIC criterion (lower panel). The results are displayed only for the horizons $h = 1$ and $h = 12$, with other results available upon request.

with the assertion of Mark (1995), and many others, that predictive ability of exchange rates at short horizons is likely to be low.

The results of model selection performed by the standard BIC criterion, seen in the lower panels of Tables 3 and 4, show that much larger models are selected for all countries, both pre- and post-Euro. This result was expected from the theory in section 3 which showed the BIC to be inconsistent, and confirms that it is unwise to rely on these standard selection criteria in empirical studies using the idiosyncratic components estimated from a factor model.

In order to explore the sensitivity of these results we ran some additional results, all of which can be found in the Online Appendix to this paper. We firstly repeated the exercise using the

$HQICM$ criterion of Groen and Kapetanios (2013) rather than the HQ_3 criterion of this paper. Since both criteria are consistent, they should select a similar number of idiosyncratic components. The results in the Online Appendix, indeed, show that very similar results obtain when using the $HQICM$ criterion, with slightly more conservative selection in the post-Euro sub-sample.

Another set of results looked at whether the spillover effects had more of an impact during recessions. Since we use a panel of different countries, rather than using individual countries' recession dates, we used the dates for global recessions identified by the International Monetary Fund (IMF), the last four of which have been in 1975, 1982, 1991 and 2009.[14] These dates are used to create a global dummy variable D_t, equal to 1 in global recession and 0 otherwise, to augment Equation (15) as follows:

$$ s_{i,t+h} - s_{it} = \mu_1 + \mu_2 D_t + \alpha_{1i}\widehat{u}_{it} + \alpha_{2i}\widehat{u}_{it}D_t + \sum_{j\neq i}\alpha_{1j}\widehat{u}_{jt} + \sum_{j\neq i}\alpha_{2j}\widehat{u}_{jt}D_t + \varepsilon_{i,t+h} \qquad (17) $$

The results in the Online Appendix show that the baseline results from Tables 3 and 4 do not change when we search over the new recession interaction terms. This firstly indicates that there does not appear to be an additional spillover effect during global recession and, secondly, that the new information criteria are robust to the inclusion of additional variables. On the other hand, the inconsistent BIC criterion continues to overfit the model and, indeed, selects some of the new recession interaction terms for some countries. This highlights that the standard BIC is not a reliable model selection device in these circumstances.

6.4 Pseudo Out-of-Sample Results

In order to assess the out-of-sample performance of the models in Equations (15) and (16), we undertake a pseudo out-of-sample forecasting experiment. For the pre- and post-Euro sub-samples, we denote the total sample size as $T+h$ and split the sample into an 'in-sample' and 'out-of-sample' portion $T = R + P - 1$, having lagged the regressors h periods for the direct forecasting scheme. We proceed to form P out-of-sample forecasts by rolling through the sample with a rolling window length R, starting with the first R observations of the sub-sample. In every rolling window, we re-estimate the factors and idiosyncratic components from the first-stage model, selecting the number of factors each time by the BIC_3 criterion of Bai and Ng (2002). Then in the second-stage we perform model selection using the HQ_3 criterion and estimate the parameters of the model by OLS before making the h-step ahead forecast.

This pseudo out-of-sample procedure yields a string of P forecast errors which, for the larger model in Equation (16), can be written for each country i as:

$$ \widehat{\varepsilon}_{i,t+h} = (s_{i,t+h} - s_{it}) - \left(\widehat{\mu}_t + \widehat{\alpha}_{it}\widehat{u}_{it}^{(t)} + \sum_{j\neq i}\widehat{\alpha}_{jt}\widehat{u}_{jt}^{(t)} + \widehat{\beta}_t\left(s_{it} - z_{it}\right) \right) \qquad (18) $$

---

[14]See Box 1.1 of the April 2009 World Economic Outlook: http://www.imf.org/external/pubs/ft/weo/2009/01/pdf/text.pdf [Last accessed: 09/09/2016].

for all $t = R, ..., T$, where the estimated parameters $\widehat{\mu}_t, \widehat{\beta}_t, \widehat{\alpha}_{it}, \widehat{\alpha}_{jt}$ are the OLS estimators using the rolling window of data from $t - R + 1, ..., t$, and the idiosyncratic errors $\widehat{u}_{it}^{(t)}$ and spillovers $\widehat{u}_{jt}^{(t)}$ are superscripted by t as they have also been estimated by Principal Components using the same rolling window of the data.

The competitor model is the no-change forecast, which gives rise to the P forecast errors:

$$\widehat{\varepsilon}_{i,t+h}^{NC} = (s_{i,t+h} - s_{it}) - 0 \tag{19}$$

for all $t = R, ..., T$. For a country i, we will compare the two sets of forecasts in the expressions (18) and (19) using the Relative Root Mean Squared Forecast Error (RMSFE) statistic:

$$RMSFE_i = \frac{\frac{1}{P} \sum_{t=R}^{T} \sqrt{\widehat{\varepsilon}_{i,t+h}^2}}{\frac{1}{P} \sum_{t=R}^{T} \sqrt{(\widehat{\varepsilon}_{i,t+h}^{NC})^2}} \tag{20}$$

Therefore, a value greater than 1 implies that the no-change benchmark outperforms the exchange rate model, whereas a value less than 1 implies that the exchange rate model improves over the no-change benchmark. In the results, rather than present individual statistics by country and model, we instead look at the median RMSFE statistic across countries, in a similar way to Engel et al. (2015).[15] We will check whether the root MSFE of the model is significantly different from the no-change benchmark using the conditional predictive ability testing approach of Giacomini and White (2006). This is appropriate to use here as it allows the comparison of 'forecasting methods' and not 'forecasting models,' whereas the standard Diebold-Mariano test would require all the models to have a fixed number of parameters in every rolling window.

Table 5 displays the results for four versions of the model in Equation (16) relative to the no-change benchmark. The first version uses only the domestic idiosyncratic component and omits the PPP and spillover terms by setting $\beta = 0$ and all $\alpha_j = 0$. The second version omits the PPP term by setting $\beta = 0$, thus corresponding to the model in Equation (15). The third omits the spillover effects by setting all $\alpha_j = 0$, and the final version is the full unrestricted model in Equation (16).

The results in Table 5 to a large extent confirm the results of other studies: that it is very difficult to beat the no-change model in out-of-sample prediction. The median RMSFE is above 1 in most cases, except for at the largest forecast horizons. This is consistent with the findings of Mark (1995), that predictive ability should increase with horizon. On the other hand, purely from a qualitative point of view, we can see that the addition of spillover effects tends to reduce the median RMSFE in many cases, both when they are added to the "$\widehat{u}_{it}$ only" model and also when added to the "$\widehat{u}_{it}$ + PPP" model. In the pre-Euro sub-sample, the best model in terms of the #(RMSFE<1) statistic is the "$\widehat{u}_{it}$ + $\widehat{u}_{jt}$" model for horizons up to $h = 12$. At the longest horizons, however, the models including the PPP variable appear to dominate, which indicates that the macroeconomic fundamentals have more forecasting power at longer horizons. However,

---

[15]Results for the individual countries are available on request from the author.

Table 5: Pseudo Out-of-Sample Forecasting Results.

		Pre-Euro						
		$h=1$	$h=3$	$h=6$	$h=9$	$h=12$	$h=18$	$h=24$
$\widehat{u}_i$ only	Median RMSFE	1.03	1.12	1.14	1.07	1.09	1.13	1.19
	#(RMSFE<1)	4	2	2	3	3	0	0
$\widehat{u}_i + \widehat{u}_j$	Median RMSFE	1.07	1.01	1.10	0.99	1.05	1.14	1.20
	#(RMSFE<1)	6	7	3	8	3	0	0
	Median $\widehat{m}$	0.23	0.64	0.22	0.20	0.08	0.07	0.18
$\widehat{u}_i + $ PPP	Median RMSFE	1.04	1.24	1.39	1.30	1.44	1.08	0.95
	#(RMSFE<1)	2	0	0	1	2	6	10
$\widehat{u}_i + $ PPP $ + \widehat{u}_j$	Median RMSFE	1.09	1.13	1.27	1.22	1.30	1.08	0.92
	#(RMSFE<1)	4	1	1	1	1	6	9
	Median $\widehat{m}$	0.27	1.14	0.65	0.42	0.51	0.76	0.83

		Post-Euro						
		$h=1$	$h=3$	$h=6$	$h=9$	$h=12$	$h=18$	$h=24$
$\widehat{u}_i$ only	Median RMSFE	1.05	1.10	1.15	1.10	1.14	1.31	1.40
	#(RMSFE<1)	3	1	1	0	0	1	1
$\widehat{u}_i + \widehat{u}_j$	Median RMSFE	1.05	1.09	1.10	1.09	1.15	1.25	1.36
	#(RMSFE<1)	3	0	1	0	0	1	1
	Median $\widehat{m}$	0.00	0.03	0.16	0.19	0.02	0.18	0.17
$\widehat{u}_i + $ PPP	Median RMSFE	1.04	1.10	1.12	1.14	1.21	1.10	1.04
	#(RMSFE<1)	3	1	0	0	0	2	3
$\widehat{u}_i + $ PPP $ + \widehat{u}_j$	Median RMSFE	1.04	1.10	1.12	1.04	1.02	1.01	1.12
	#(RMSFE<1)	3	0	2	4	3	4	2
	Median $\widehat{m}$	0.00	0.05	0.45	0.31	0.22	0.71	0.75

Notes: The Median RMSFE statistic is the median of the individual countries' relative mean squared forecast error statistic of the model relative to the no-change benchmark. The median is taken over $N = 17$ countries in the Pre-Euro panel and $N = 10$ countries for post-Euro. The #(RMSFE<1) statistic counts the number of countries for which the RMSFE is less than 1, where the model has lower MSFE than the no-change benchmark.

the evidence from the Giacomini and White (2006) test somewhat diminishes these results, as only a handful of results are statistically significantly in favour of the exchange rate model over the no-change benchmark at the 10% level, and no results are significant at the 5% level.

Purely from a model selection point of view, which is the main purpose of this paper, we can see that the HQ_3 criterion selects a small although non-trivial amount of spillover effects on average. The Median $\widehat{m}$ statistic takes the median of all countries' number of selected spillover effects, averaged over all pseudo out-of-sample observations. From this we can see that the number of chosen idiosyncratic components tends to increase over the forecast horizon, indicating that the

explanatory power of these variables tends to be for longer-term prediction. The numbers are quite conservative, typically showing that between 0 and 1 idiosyncratic spillover terms are selected on average.[16] In a similar way to the in-sample results, there are fewer spillover terms selected in the post-Euro sub-sample relative to pre-Euro.

7   Conclusion

In this paper we have proposed information criteria for performing model selection in regressions involving both estimated factors and idiosyncratic components. Our main contribution is a theorem on selection consistency which shows that new conditions are required for consistent model selection, which render existing model selection criteria inconsistent in this set-up. Specifically, using Principal Components, we show that estimation error in the idiosyncratic component vanishes at a rate $\min\left\{\sqrt{T}, N\right\}$, whereas factor estimation error vanishes at a rate $\min\{T, N\}$. This implies that even criteria such as those of Bai and Ng (2009) and Groen and Kapetanios (2013), which account for factor estimation error, are inconsistent in cases where N is of larger order than $\sqrt{T}$. Furthermore, standard information criteria such as the BIC, which is used in many empirical forecasting applications, is inconsistent regardless of the sample size. These should not be used to specify models involving factors and idiosyncratic components.

We therefore propose a new set of information criteria which satisfy the conditions required for consistency and take the estimation error in the idiosyncratic component into account. We show that these new criteria perform well in Monte Carlo simulations, relative to the existing information criteria which severely overfit the models in some or all configurations of N and T. We illustrate these model selection methods with an empirical application to forecasting exchange rates, extending the recent model of Engel et al. (2015) to allow for exchange rate spillover effects. We find that our methods select a non-zero amount of spillover effects, even in a challenging predictive environment when the models do not perform much better than a no-change benchmark. Future work can look at penalized LASSO-type regression for selection of these models. This was proposed without formal justification by the empirical study of Luciani (2014). Formal results using penalized regressions would provide a useful alternative to the information criteria proposed in this paper.

References

Amengual, D. and M. W. Watson (2007). Consistent estimation of the number of dynamic factors in a large n and t panel. *Journal of Business and Economic Statistics 25*(1), 91–96.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

---

[16]We also conducted an analysis of the number of idiosyncratic components selected over time. While there was a small amount of time-variability in the number of selected components, these results were omitted as there were no clear patterns which were common across all countries.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica 74*(4), 1133–1150.

Bai, J. and S. Ng (2009). Boosting diffusion indices. *Journal of Applied Econometrics 24*(4), 607–629.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132*(1), 169–194.

Brownlees, C. T., E. Nualart, and Y. Sun (2015). Realized networks. *Available at SSRN 2506703*.

Carrasco, M. and B. Rossi (2016). In-sample Inference and Forecasting in Misspecified Factor Models. *Journal of Business & Economic Statistics* (Forthcoming).

Castle, J. L., M. P. Clements, and D. F. Hendry (2013). Forecasting by factors, by variables, by both or neither? *Journal of Econometrics 177*(2), 305–319.

Cheng, X. and B. E. Hansen (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics 186*(2), 280–293.

De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting Using a Large Number of Predictors: Is Bayesian Shrinkage a Valid Alternative to Principal Components? *Journal of Econometrics 146*(2), 318–328.

Djogbenou, A. (2016). Model Selection in Factor-Augmented Regressions with Estimated Factors. *Mimeo*.

Doz, C., D. Giannone, and L. Reichlin (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics 164*(1), 188–205.

Doz, C., D. Giannone, and L. Reichlin (2012). A quasi-maximum likelihood approach for large, approximate dynamic factor models. *The review of economics and statistics 94*(4), 1014–1024.

Engel, C., K. West, and N. Mark (2015). Factor model forecasts of exchange rates. *Econometric Reviews 34*(1), 32–55.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Gonçalves, S. and B. Perron (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics 182*(1), 156–173.

Groen, J. J. J. and G. Kapetanios (2013). Model selection criteria for factor-augmented regressions. *Oxford Bulletin of Economics and Statistics 75*(1), 37–63.

Hallin, M. and R. Liška (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association 102*(478), 603–617.

Hansen, B. E. (2007). Least Squares Model Averaging. *Econometrica 75*(4), 1175–1189.

Kilian, L. (1999). Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions? *Journal of Applied Econometrics 14*(5), 491–510.

Luciani, M. (2014). Forecasting with approximate dynamic factor models: the role of non-pervasive shocks. *International Journal of Forecasting 30*(1), 20–29.

Mark, N. (1995). Exchange rates and fundamentals: evidence on long-horizon predictability. *The American Economic Review 85*(1), 201–218.

McCracken, M. W. and S. G. Sapp (2005). Evaluating the predictability of exchange rates using long-horizon regressions: Mind your p's and q's! *Journal of Money, Credit and Banking 37*(3), 473–494.

Meese, R. and K. Rogoff (1983a). The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification? In J. A. Frenkel (Ed.), *Exchange Rates and International Macroeconomics*, pp. 67–112. Chicago: University of Chicago Press.

Meese, R. A. and K. Rogoff (1983b). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics 14*(1), 3.

Ng, S. (2013). Variable selection in predictive regressions. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2. North-Holland.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics 92*(4), 1004–1016.

Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature 51*(4), 1063–1119.

Stock, J. and M. Watson (2011). Dynamic factor models. In M. P. Clements and D. F. Hendry (Eds.), *The Oxford handbook of economic forecasting*, pp. 35–60. New York: Oxford University Press.

Stock, J. H. and M. W. Watson (1999). Forecasting inflation. *Journal of Monetary Economics 44*(2), 293–335.

Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics 20*(2), 147–162.
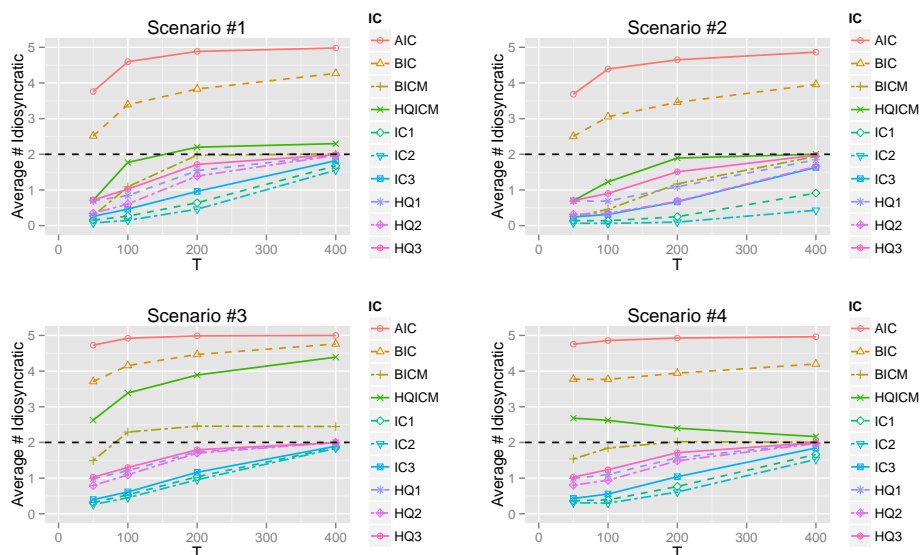
8 Figures in the Text

Figure 1: Average number of selected idiosyncratic components $(\widehat{m}^0)$ for different information criteria over 1,000 Monte Carlo replications when the true $r = 1$ and $m^0 = 1$
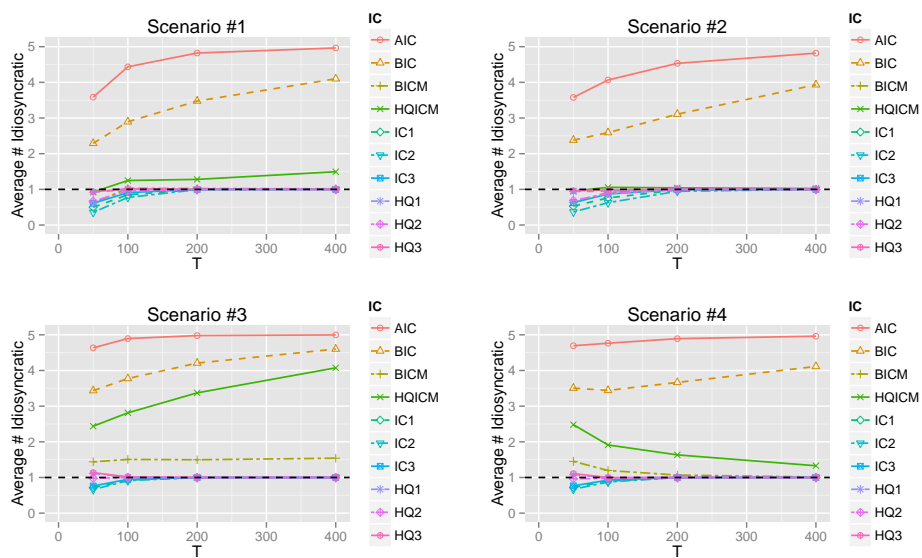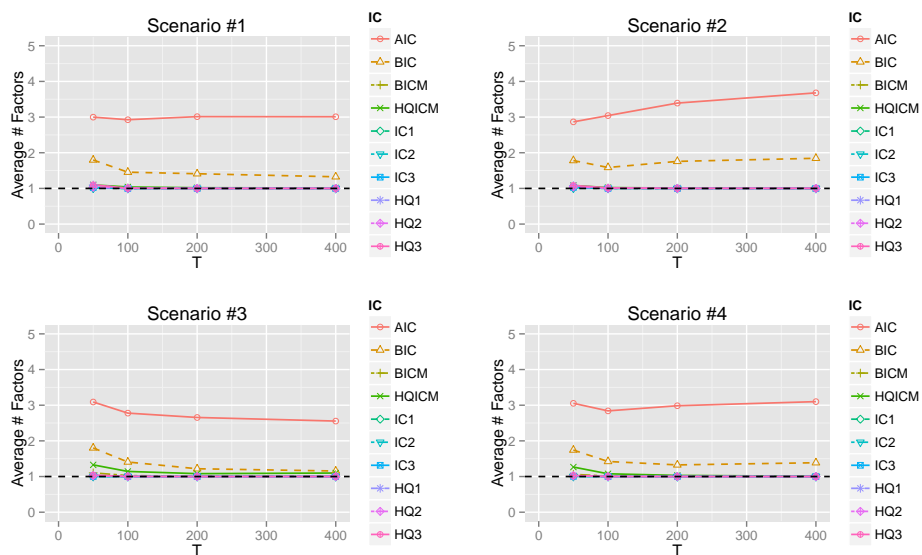

Notes: The horizontal dashed line represents the true number of idiosyncratic components $m^0 = 1$. Scenarios 1-4 are described in Section 5.1 and each of the information criteria are described in Section 4.

Figure 2: Average number of selected idiosyncratic components $(\widehat{m}^0)$ for different information criteria over 1,000 Monte Carlo replications when the true $r = 1$ and $m^0 = 2$
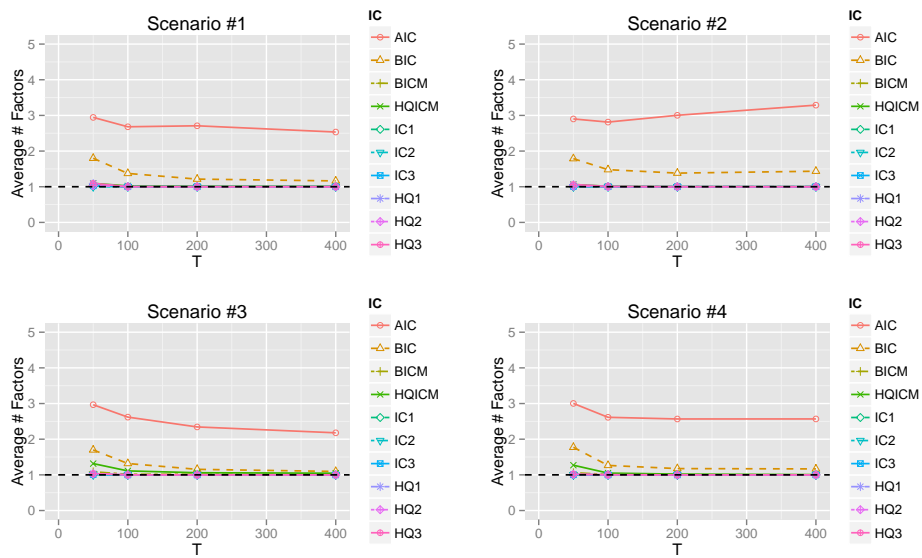

Notes: The horizontal dashed line represents the true number of idiosyncratic components $m^0 = 2$. Scenarios 1-4 are described in Section 5.1 and each of the information criteria are described in Section 4.

Figure 3: Average number of selected factors $(\widehat{r})$ for different information criteria over 1,000 Monte Carlo replications when the true $r = 1$ and $m^0 = 1$

Notes: The horizontal dashed line represents the true number of factors $r = 1$. Scenarios 1-4 are described in Section 5.1 and each of the information criteria are described in Section 4.

Figure 4: Average number of selected factors ($\widehat{r}$) for different information criteria over 1,000 Monte Carlo replications when the true $r = 1$ and $m^0 = 2$

Notes: The horizontal dashed line represents the true number of factors $r = 1$. Scenarios 1-4 are described in Section 5.1 and each of the information criteria are described in Section 4.