Econometric Theory, **24**, 2008, 553–579. Printed in the United States of America.
DOI: 10.1017/S0266466608080237

USING MACRO DATA TO OBTAIN BETTER MICRO FORECASTS

Jan R. Magnus and Andrey L. Vasnev
Tilburg University, The Netherlands

We consider the problem of combining forecasts from two different levels (called "macro" and "micro"), where we have access to the forecasts and their precisions but not to the full data set. We develop a theoretical framework and provide Monte Carlo evidence in the cases of both perfect and imperfect aggregation. Our proposed procedure is simple and robust. We also extend the procedure to time series and propose a forecast model for the European zero rates, combining quarterly and monthly observations. We show that forecast accuracy is improved at both levels.

1. INTRODUCTION

Data are often available at different levels or at different frequencies. In cross sections, relevant data may be available at a macro level (industries) and at a micro level (firms), where the different levels give rise to different models, estimates, and forecasts. The models are often estimated by separate offices and statisticians, although they are clearly related to each other. For example, Marcellino, Stock, and Watson (2003) forecast aggregate Euro-wide inflation (macro) and country-specific inflation indexes (micro), and Hendry and Hubrich (2005) study aggregate forecasts of Euro-area inflation (macro) and five subcomponents: unprocessed food, processed food, industrial goods, energy, and services prices (micro). In the empirical part of this paper we shall consider an application in the time domain: quarterly ("macro") and monthly ("micro") forecasts for the Euro yield curve.

If all data were available to us at both levels, combining data would lead to efficiency gains. We shall assume, however, that we have no access to the underlying data but that we do have information on the one-period-ahead forecasts, both at the micro level and at the macro level. We now wish to combine these forecasts to obtain better forecasts at both levels, taking into account our knowledge of the relationship between the two levels.

In a linear world, the models at the two levels would be analogous,[1] but in a nonlinear world the models could (and probably should) be quite different.[2]

We are grateful to the editors of this special issue and to two referees for their constructive comments, which greatly helped improve the paper. Earlier versions of this paper were presented at the NAKE workshop in Amsterdam, at Tilburg University, and at ESEM 2006 in Vienna. We thank the participants for their useful comments. In addition, we thank Steffan Berridge, John Einmahl, and Hans Schumacher for their constructive comments. Address correspondence to Jan R. Magnus, Department of Econometrics and Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; e-mail: magnus@uvt.nl.

 0266-4666/08 $15.00

Even so, the two levels are related to each other, and therefore efficiency gains can be achieved by using information at both levels.

There is an extensive literature on combining models or forecasts at the *same* level. Bates and Granger (1969) show that optimal forecast combination reduces the variance. More recently, Leung and Barron (2006) use risk terminology and provide evidence that the performance of a mixture is superior to the best of the individual models. Hendry and Clements (2004) provide an extensive survey of the literature and list the advantages one can expect from pooling forecasts. Yang (2004) gives theoretical results for combining procedures and derives risk bounds to quantify the potential gain and the "price" of linearly combining forecasts. Other useful surveys are provided by Clemen (1989) and Diebold and Lopez (1996). Bayesian model averaging (Lancaster, 2004, p. 101) has gained popularity because of its natural interpretation and good performance in practice; see also Raftery, Madigan, and Hoeting (1997) and Hoeting, Madigan, Raftery, and Volinsky (1999).

The literature on the combination of models and forecasts at *different* levels is much less extensive. Imbens and Lancaster (1994) offer an excellent solution, assuming availability of all micro data and macro information in the form of moment restrictions. Employing the generalized method of moments (GMM) methodology, they thus obtain more precise estimates at the micro level. Their analysis is constrained to one time period only and to the availability of all micro data. Magnus, van Tongeren, and de Vos (2000) offer a solution from a Bayesian perspective in the context of national accounts estimation. More recently, Hendry and Hubrich (2005) use a simple intuitive vector autoregression (VAR) example and demonstrate the predictability of the aggregated error from a disaggregated information set, thus motivating the combination of data sets.

The contribution of this paper is not on model averaging but is on the combination of information at different levels of aggregation. Going from the detailed (micro) level to the more general (macro) level can usually be achieved through summation or averaging. Our main objective, therefore, concerns the opposite direction, that is, using macro data to improve micro forecasts. This is the first way in which this paper differs from most of the literature on combining models and forecasts: we consider models at different levels. The second is that we do not assume that all data are available to us. We propose a two-level hierarchical model, which allows us to update the micro forecast on the basis of the macro forecast. The proposed method has a natural interpretation and shows good performance in Monte Carlo simulations.

The paper is organized as follows. In Sections 2 and 3 we provide the theoretical framework and derive the proposed forecasts. In Section 4 we present Monte Carlo evidence in a simple setup (the base scenario), and in Section 5 we consider more realistic situations, involving measurement error and other deviations from the base scenario. Section 6 extends the procedure to time series. An empirical application in Section 7 shows the benefits of combining monthly and quarterly forecasts for the Euro yield curve. Our conclusion in Section 8 is that a simple and robust forecast combination rule performs uniformly well and has, in addition, a minimum mean squared error, a least squares, and also a Bayesian interpretation. Appendix A discusses a somewhat counterintuitive result about linear combinations, whereas Appendix B summarizes the data used in the empirical example.

2. A TWO-LEVEL HIERARCHICAL MODEL

We assume that information is available at two levels: the macro level and the micro level. The macro level provides aggregate information, whereas the micro level provides disaggregate information, thus giving more detail for the components of the macro level. The macro level is indexed by $t = 1, \ldots, T$; the micro level, in addition, by $j = 1, \ldots, p$.

For the two levels we may have different models and different data sources. At the macro level we assume that data are available on the variable of interest η_t together with $\ell \times 1$ vectors of exogenous explanatory variables z_t. For simplicity, we assume a linear model

$$\eta_t = z'_{t-1}\gamma + u_t, \tag{1}$$

where $\{u_t\}$ is unobservable independent and identically distributed (i.i.d.) random noise with mean zero and variance σ_u^2 and z_{t-1} is revealed at the end of period $t-1$. If $\hat{\gamma}$ denotes the ordinary least squares (OLS) estimator of γ based on T observations, then the macro forecast is given by $\hat{\eta}^{(1)}_{T+1} = z'_T\hat{\gamma}$.

The data at the more detailed micro level consist of the vector of variables of interest $y_t := (y_{t,1}, \ldots, y_{t,p})'$ together with $p \times k$ matrices of exogenous explanatory variables X_t. We assume again a linear model

$$y_t = X_{t-1}\beta + v_t, \tag{2}$$

where the vectors $\{v_t\}$ are unobservable i.i.d. random noise with mean zero and variance matrix Σ_v and X_{t-1} is revealed at the end of period $t-1$. If $\hat{\beta}$ denotes the generalized least squares (GLS) estimator of β based on T observations, then the micro forecast is given by $\hat{y}^{(2)}_{T+1} = X_T\hat{\beta}$.

The data for the macro model are compiled in a macro office, and γ is estimated by a macro statistician. The data for the micro model are compiled in a micro office, and β is estimated by a micro statistician. The two offices and the two statisticians may or may not talk to each other. Thus, some of the information used by one statistician may also be used by the other but not necessarily so. The two levels are related to each other, even though the macro and micro statisticians do not utilize this relationship. In the background lives the linear constraint

$$\eta_t = a'y_t,$$

where a is a known $p \times 1$ vector. For example, if $a = \imath_p$, the vector of ones, then the macro observations are the sum of the micro observations, and if $a = \imath_p/p$, then the macro observations are the average of the micro observations. As a result, the two levels may be (and probably will be) correlated, but we shall assume that this correlation exists only within the same time period. Hence, $\text{cov}(u_t, v_t)$ may be nonzero but $\text{cov}(u_s, v_t) = 0$ for $s \neq t$. We emphasize that we do not require the linear constraint $\eta_t = a'y_t$ to hold for every t. This would be the case of perfect model aggregation, which is rare in practice. If one, nevertheless, encounters this case, then the micro model contains the best information for both (micro and macro) levels.

The macro statistician provides a macro forecast $\hat{\eta}_{T+1}^{(1)}$ with an estimated variance $\hat{\sigma}_1^2$, and the micro statistician provides a micro forecast $\hat{y}_{T+1}^{(2)}$ with an estimated variance matrix $\hat{\Sigma}$. (We will see later that only the *relative* precisions matter.) Our task is to combine the two forecasts, making use of all available information. In particular, we wish to obtain optimal one-period-ahead forecasts $\hat{y}_{T+1}$ (micro) and $\hat{\eta}_{T+1}$ (macro), satisfying the micro-macro relationship

$$\hat{\eta}_{T+1} = a'\hat{y}_{T+1}. \tag{3}$$

Note that we do not assume that $\eta_t = a'y_t$ holds at $t = T + 1$; only that the forecast satisfies (3). Note also that our assumption on the availability of data rules out the possibility of estimating of (1) and (2) jointly. We emphasize that the linearity assumptions in (1)–(3) are made for simplicity only and do not essentially affect our analysis.

We have two forecasts of η_{T+1}, one based only on macro data and one based only on micro data:

$$\hat{\eta}_{T+1}^{(1)} = z_T'\hat{\gamma} \quad \text{and} \quad \hat{\eta}_{T+1}^{(2)} := a'\hat{y}_{T+1}^{(2)} = a'X_T\hat{\beta}.$$

We wish to improve the forecasts at both levels, and we propose the combined macro forecast as the linear combination

$$\hat{\eta}_{T+1} = \alpha\hat{\eta}_{T+1}^{(1)} + (1 - \alpha)\hat{\eta}_{T+1}^{(2)}, \tag{4}$$

in the spirit of Bates and Granger (1969). Conditional on the choice of α, this provides a common-sense improvement of the macro forecast. We call (4) a linear combination rather than a weighted average, because α may not lie between zero and one if the correlation between $\hat{\eta}_{T+1}^{(1)}$ and $\hat{\eta}_{T+1}^{(2)}$ is high and positive. This is somewhat counterintuitive, and Appendix A therefore provides a justification and explanation.

Two questions now arise. First, how do we choose α? Second, how do we update the micro forecast, given the improvement of the macro forecast? We shall discuss the second question first.

2.1. Updating the Micro Forecast

We assume first that α is known and that the macro forecast is given by (4). Given the improvement at the macro level, we now update the micro forecast.

Letting $\Sigma := \text{var}(\hat{y}^{(2)}_{T+1}) = X_T \text{var}(\hat{\beta}) X'_T$, this is achieved by choosing y to maximize

$$\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(y - \hat{y}^{(2)}_{T+1})'\Sigma^{-1}(y - \hat{y}^{(2)}_{T+1})\right\}$$

subject to

$$a'y = \hat{\eta}_{T+1}. \tag{5}$$

Under normality, the solution is the point with the highest probability in the micro model satisfying the macro restriction. In the two-dimensional case, Figure 1 illustrates the intuition. We have a point estimate (c_1, c_2) and the associated probability distribution, and we want the sum of the two estimates to be equal to c'. To include this additional information, we choose the ellipse (constant-probability contour) that is tangent to the line $c'_1 + c'_2 = c'$. The tangent point satisfies the restriction and is the most probable point given the probability distribution. In the special case where the forecasts c_1 and c_2 are not correlated, they will be adjusted proportionally to their variances. In the special case where the variance of c_1 is zero, all adjustment will be absorbed by c_2.

The normality assumption is of course a simplification. We use it here because of its central role as an approximating distribution and because it provides explicit and intuitively clear results. Our procedure can be generalized to other densities, but a closed-form solution will then typically not be available and numerical optimization has to be performed.

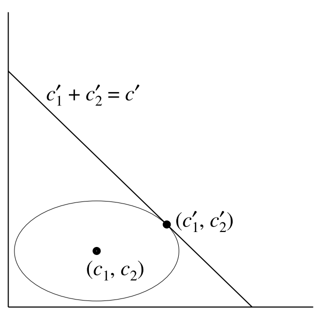

FIGURE 1. Update procedure in the two-dimensional case.

The solution to (5) is given by

$$\hat{y}_{T+1} = \hat{y}^{(2)}_{T+1} + \frac{\alpha(\hat{\eta}^{(1)}_{T+1} - \hat{\eta}^{(2)}_{T+1})}{a'\Sigma a}\,\Sigma a, \tag{6}$$

which can be calculated if we replace Σ by $\hat{\Sigma}$ and if we know α. Note that the forecast (6) is biased, because the expectations of $\hat{\eta}^{(1)}_{T+1}$ and $\hat{\eta}^{(2)}_{T+1}$ are, in general, not equal. Thus we obtain a macro forecast $\hat{\eta}_{T+1}$ and a micro forecast $\hat{y}_{T+1}$, satisfying the restriction $\hat{\eta}_{T+1} = a'\hat{y}_{T+1}$ and taking into account information at both levels.

2.2. The Choice of α

We next address the question how to choose α. We shall employ the mean squared error criterion as our guiding principle, although this criterion has some drawbacks (Clements and Hendry, 1993) and some other criteria have recently been proposed by Diebold and Kilian (2001), Inoue and Kilian (2006), and Yang (2004).

The mean squared error of the forecast (6) is given by

$$\begin{aligned} \text{MSE}(\hat{y}_{T+1}) &= \text{E}(y_{T+1} - \hat{y}_{T+1})(y_{T+1} - \hat{y}_{T+1})' \\ &= \text{MSE}(\hat{y}^{(2)}_{T+1}) + \frac{(\alpha q - c)(\alpha q - c)' - cc'}{\text{E}(\delta^2)}, \end{aligned} \tag{7}$$

where

$$q := \frac{\text{E}(\delta^2)}{a'\Sigma a}\,\Sigma a, \qquad c := \text{E}(\delta e)$$

and

$$\delta := \hat{\eta}^{(1)}_{T+1} - \hat{\eta}^{(2)}_{T+1} = z'_T\hat{\gamma} - a'X_T\hat{\beta},$$

$$e := y_{T+1} - \hat{y}^{(2)}_{T+1} = X_T(\beta - \hat{\beta}) + v_{T+1}.$$

Hence, $\text{MSE}(\hat{y}_{T+1})$—considered as a function of α—is “small” if and only if $(\alpha q - c)(\alpha q - c)'$ is “small.” The latter matrix has rank one, and it possesses one positive eigenvalue $(\alpha q - c)'(\alpha q - c)$, which is minimized for

$$\alpha_1 = \frac{q'c}{q'q} = \frac{(a'\Sigma(\text{E}\delta e))(a'\Sigma a)}{(\text{E}\delta^2)(a'\Sigma^2 a)}.$$

Choosing $\alpha = \alpha_1$ thus minimizes the trace of $\text{MSE}(\hat{y}_{T+1})$.[3] Using the facts that

$$\text{E}(y_{T+1} - X_T\hat{\beta}) = 0 \quad \text{and} \quad \text{cov}(\delta, y_{T+1}) = 0,$$

we obtain

$$\mathrm{E}(\delta e) = \Sigma a - \pi \quad \text{and} \quad \mathrm{E}(\delta^2) = \sigma_1^2 + \sigma_2^2 - 2a'\pi + (z_T'\gamma - a'X_T\beta)^2,$$

and hence

$$\alpha_1 = \frac{(a'\Sigma(\Sigma a - \pi))\sigma_2^2}{((z_T'\gamma - a'X_T\beta)^2 + \sigma_1^2 + \sigma_2^2 - 2a'\pi)(a'\Sigma^2 a)}, \tag{8}$$

where

$$\sigma_1^2 := \mathrm{var}(\hat{\eta}_{T+1}^{(1)}) = z_T'\mathrm{var}(\hat{\gamma})z_T, \quad \sigma_2^2 := \mathrm{var}(\hat{\eta}_{T+1}^{(2)}) = a'\Sigma a$$

and

$$\pi := \mathrm{cov}(\hat{y}_{T+1}^{(2)}, \hat{\eta}_{T+1}^{(1)}) = X_T\mathrm{cov}(\hat{\beta}, \hat{\gamma})z_T.$$

Because information on micro and macro forecasts reaches us separately, we do not have a direct estimate of π. We can, however, estimate π from a history of forecasts $\{\hat{y}_{T+1}^{(2)}, \hat{\eta}_{T+1}^{(1)}\}$, if such a history is available. Quite often, such a history will not be available. Interestingly, our simulations (to be discussed later) indicate that the assumption $\pi = 0$ is a reasonable, often optimal, assumption to make.

If the two processes $\{u_t\}$ and $\{v_t\}$ are uncorrelated, we have $\pi = 0$, and hence

$$\alpha_2 = \frac{\sigma_2^2}{(z_T'\gamma - a'X_T\beta)^2 + \sigma_1^2 + \sigma_2^2}, \tag{9}$$

which is a number between zero and one. For small values of the components of π, α_1 also lies between zero and one. But if the correlation between $\hat{\beta}$ and $\hat{\gamma}$ is positive and high, then α_1 does not necessarily lie between zero and one.

To compare the various combined forecasts we consider the trace of $\mathrm{MSE}(\hat{y}_{T+1})$, which we denote generically by τ^2. For $\alpha = \alpha_1$ we obtain from (7)

$$\begin{aligned}\tau_1^2 &= \mathrm{tr}(\mathrm{MSE}(\hat{y}_{T+1}^{(2)})) + \frac{(\alpha_1 q - c)'(\alpha_1 q - c) - c'c}{\mathrm{E}(\delta^2)} \\ &= \mathrm{tr}(\mathrm{MSE}(\hat{y}_{T+1}^{(2)})) - \frac{(a'\Sigma(\Sigma a - \pi))^2}{((z_T'\gamma - a'X_T\beta)^2 + \sigma_1^2 + \sigma_2^2 - 2a'\pi)(a'\Sigma^2 a)}\end{aligned} \tag{10}$$

and for $\alpha = \alpha_2$

$$\tau_2^2 = \mathrm{tr}(\mathrm{MSE}(\hat{y}_{T+1}^{(2)})) - \frac{a'\Sigma^2 a}{(z_T'\gamma - a'X_T\beta)^2 + \sigma_1^2 + \sigma_2^2} \geq \tau_1^2. \tag{11}$$

3. A DIFFERENT ROUTE

The method of the previous section can be summarized as follows. Given a macro forecast $\hat{\eta}_{T+1}^{(1)}$ and a micro forecast $\hat{y}_{T+1}^{(2)}$, we minimize the quadratic form

$$(y - \hat{y}_{T+1}^{(2)})' \Sigma^{-1} (y - \hat{y}_{T+1}^{(2)})$$

subject to the restriction

$$a'y = \alpha \hat{\eta}_{T+1}^{(1)} + (1-\alpha) a' \hat{y}_{T+1}^{(2)}.$$

This produces a forecast $\hat{y}_{T+1}(\alpha)$. We then choose α such that the mean squared error of $\hat{y}_{T+1}(\alpha)$ is minimized.

We now describe a second possible method, based on the idea that we should choose y such that $a'y$ is close to $\hat{\eta}_{T+1}^{(1)}$ and that y is close to $\hat{y}_{T+1}^{(2)}$. Thus we minimize

$$\begin{pmatrix} a'y - \hat{\eta}_{T+1}^{(1)} \\ y - \hat{y}_{T+1}^{(2)} \end{pmatrix}' \Omega^{-1} \begin{pmatrix} a'y - \hat{\eta}_{T+1}^{(1)} \\ y - \hat{y}_{T+1}^{(2)} \end{pmatrix},$$

where

$$\Omega := \begin{pmatrix} \sigma_1^2 & \pi' \\ \pi & \Sigma \end{pmatrix}.$$

We can also phrase this problem as the regression problem

$$\begin{pmatrix} \hat{\eta}_{T+1}^{(1)} \\ \hat{y}_{T+1}^{(2)} \end{pmatrix} = \begin{pmatrix} a' \\ I_p \end{pmatrix} y + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix},$$

with

$$\mathrm{E}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = 0, \qquad \mathrm{var}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \Omega.$$

This formulation implies that $\mathrm{E}(\hat{\eta}_{T+1}^{(1)}) = \mathrm{E}(a'\hat{y}_{T+1}^{(2)})$, so that $\mathrm{E}(\delta) = 0$. After somewhat tedious but straightforward calculations we find the GLS solution as

$$\hat{y}_{T+1} = \hat{y}_{T+1}^{(2)} + \frac{\alpha_3 (\hat{\eta}_{T+1}^{(1)} - \hat{\eta}_{T+1}^{(2)})}{\sigma_2^2 - a'\pi} (\Sigma a - \pi), \qquad \textbf{(12)}$$

with

$$\alpha_3 := \frac{\sigma_2^2 - a'\pi}{\sigma_1^2 + \sigma_2^2 - 2a'\pi}. \qquad \textbf{(13)}$$

Note that we introduce α_3 to connect the GLS solution (12) to the previous solutions. There is, however, no optimization over α in this case. In this sense one can think about the GLS solution as optimal for both micro and macro level.

The forecast (12) is unbiased, and its variance is given by

$$\operatorname{var}(\hat{y}_{T+1}) = \Sigma - \frac{1}{\sigma_1^2 + \sigma_2^2 - 2a'\pi} (\Sigma a - \pi)(\Sigma a - \pi)',$$

so that

$$\tau_3^2 = \operatorname{tr}(\operatorname{MSE}(\hat{y}_{T+1})) = \operatorname{tr}(\Sigma) - \frac{(\Sigma a - \pi)'(\Sigma a - \pi)}{\sigma_1^2 + \sigma_2^2 - 2a'\pi}, \tag{14}$$

which should be viewed as the analogue to τ_1^2 in the previous section. There is no theoretical inequality between τ_1^2 and τ_3^2. For example, in the special case when there is no inconsistency between the models (so that $z_T'\gamma - a'X_T\beta = 0$), and the variance structure is simply $\Sigma = I_p$ and $a = \iota_p$, we find

$$\tau_1^2 = p - \frac{(p - a'\pi)^2}{(\sigma_1^2 + \sigma_2^2 - 2a'\pi)p} \quad \text{and} \quad \tau_3^2 = p - \frac{(a - \pi)'(a - \pi)}{\sigma_1^2 + \sigma_2^2 - 2a'\pi}.$$

The difference between τ_1^2 and τ_3^2 then depends on the relationship between

$$\left(\frac{a'\pi}{a'a}\right)^2 = \left(\frac{1}{p}\sum \pi_i\right)^2 \quad \text{and} \quad \frac{\pi'\pi}{a'a} = \frac{1}{p}\sum \pi_i^2,$$

and both $\tau_1^2 < \tau_3^2$ and $\tau_1^2 > \tau_3^2$ are possible.

Given the micro forecast, the macro forecast is then given by

$$\hat{\eta}_{T+1} := a'\hat{y}_{T+1} = \alpha_3 \hat{\eta}_{T+1}^{(1)} + (1 - \alpha_3)\hat{\eta}_{T+1}^{(2)}.$$

In the special case where $\pi = 0$ we obtain

$$\hat{y}_{T+1} = \hat{y}_{T+1}^{(2)} + \frac{\alpha_4(\hat{\eta}_{T+1}^{(1)} - \hat{\eta}_{T+1}^{(2)})}{\sigma_2^2} \Sigma a, \tag{15}$$

with

$$\alpha_4 := \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \tag{16}$$

The variance of this forecast is

$$\operatorname{var}(\hat{y}_{T+1}) = \Sigma - \frac{1}{\sigma_1^2 + \sigma_2^2} \Sigma a a' \Sigma, \tag{17}$$

so that

$$\tau_4^2 = \operatorname{tr}(\mathrm{MSE}(\hat{y}_{T+1})) = \operatorname{tr}(\Sigma) - \frac{a'\Sigma^2 a}{\sigma_1^2 + \sigma_2^2}. \tag{18}$$

Notice that $\tau_4^2 \leq \tau_2^2$ as a result of the fact that in the derivation of τ_4^2 we have assumed that the inconsistency parameter satisfies $\mathrm{E}(\delta) = 0$.

The result (15) can also be phrased in Bayesian terms by considering macro data together with a micro prior:

$$\hat{\eta}_{T+1}^{(1)} | y_{T+1} \sim \mathrm{N}(a' y_{T+1}, \sigma_1^2), \qquad y_{T+1} \sim \mathrm{N}_p(\hat{y}_{T+1}^{(2)}, \Sigma).$$

Using Theorem 1 of Magnus, van Tongeren, and de Vos (2000), we then obtain the mean and the variance of the posterior distribution of $y_{T+1} | \hat{\eta}_{T+1}^{(1)}$ as (15) and (17), respectively.

Finally, if more information were available than there is, we could try to incorporate macro information in the micro model, motivated by Hendry and Hubrich (2005). For example, we might assume

$$y_t = \frac{1}{1 + a'h} X_{t-1}\beta + \frac{h}{1 + a'h} z_{t-1}'\gamma + \zeta_t = \widetilde{X}_{t-1}\begin{pmatrix}\beta \\ \gamma\end{pmatrix} + \zeta_t, \tag{19}$$

where the vector h will typically be the unit vector $\imath$ and $a'h > 0$. The formulation (19) implies that $a'y_t$ is a weighted average of $a'X_{t-1}\beta$ and $z_{t-1}'\gamma$, apart from some noise $a'\zeta_t$.

We emphasize that this model requires knowledge of the separate micro and macro data sets, which we do not have. Nevertheless it is of interest to compare our feasible forecasts with some forecasts that require more data, because this provides information on the sensitivity of the feasible forecasts.

Let $\tilde{\beta}$ and $\tilde{\gamma}$ denote the OLS estimators in this model. The unrestricted estimators $\tilde{\beta}$ and $\tilde{\gamma}$ are related to the restricted estimator $\hat{\beta}$ (under the restriction $\gamma = 0$) by $\hat{\beta} = \tilde{\beta} - \Pi_1\tilde{\gamma}$ for some matrix Π_1. Hence the micro forecast is given by

$$\begin{aligned}\tilde{y}_{T+1} &= \widetilde{X}_T\begin{pmatrix}\tilde{\beta} \\ \tilde{\gamma}\end{pmatrix} = \widetilde{X}_T\begin{pmatrix}\hat{\beta} + \Pi_1\tilde{\gamma} \\ \tilde{\gamma}\end{pmatrix} \\ &= \widetilde{X}_T\begin{pmatrix}\hat{\beta} \\ 0\end{pmatrix} + \widetilde{X}_T\begin{pmatrix}\Pi_1 \\ I\end{pmatrix}\tilde{\gamma} = \hat{y}_{T+1}^{(2)} + \widetilde{X}_T\Pi\tilde{\gamma},\end{aligned} \tag{20}$$

where $\Pi' = (\Pi_1' : I)$ and its mean squared error is

$$\mathrm{MSE}(\tilde{y}_{T+1}) = \mathrm{MSE}(\hat{y}_{T+1}^{(2)}) + \widetilde{X}_T(\Pi C' + C\Pi' + \Pi\mathrm{E}(\tilde{\gamma}\tilde{\gamma}')\Pi')\widetilde{X}_T',$$

where

$$C := \begin{pmatrix} -\mathrm{cov}(\hat{\beta}, \tilde{\gamma}) \\ \gamma\gamma' \end{pmatrix}.$$

We see that $\tau_5^2 := \mathrm{tr}(\mathrm{MSE}(\tilde{y}_{T+1}))$ in this case can be larger or smaller than $\mathrm{tr}(\mathrm{MSE}(\hat{y}_{T+1}^{(2)}))$.

4. MONTE CARLO EVIDENCE: THE BASE SCENARIO

To compare the proposed forecasts and to assess the possible improvement resulting from the use of macro information, extensive simulations were performed on which we now report. First we generate five regressors:

x_1 constant: $1, 1, 1, \dots,$

x_2 time trend: $1, 2, 3, \dots,$

x_3 normal distribution: $\mathrm{E}(x_3) = 0, \mathrm{var}(x_3) = 9,$

x_4 lognormal distribution: $\mathrm{E}(\log x_4) = 0, \mathrm{var}(\log x_4) = 9,$

x_5 uniform distribution: $-2 \leq x_5 \leq 2.$

These regressors can be combined in various data sets. We consider five data sets with two regressors and five with three regressors, as follows:

	$k = 2$		$k = 3$
1	constant, time trend	6	constant, time trend, normal
2	constant, normal	7	constant, time trend, lognormal
3	constant, lognormal	8	constant, uniform, lognormal
4	uniform, normal	9	uniform, normal, lognormal
5	time trend, normal	10	time trend, normal, uniform

The number of regressors is therefore $k = 2$ or $k = 3$. Given one of these 10 data sets, we generate the $p \times k$ matrices X_{t-1} for $t = 1, \dots, T$. We consider $T = 24$ or $T = 96$ years and within each year $p = 3$ or $p = 9$ micro units.

4.1. The Data-Generation Process

For the base scenario, the micro data are generated by

$$y_t = X_{t-1}\beta + v_t, \qquad v_t \sim \mathrm{N}(0, \sigma_v^2 I_p), \tag{21}$$

where all k elements of β are equal to one. Letting $\sigma_v^2 = 1$ we take independent draws from the $\mathrm{N}(0, \sigma_v^2)$ distribution, generate the $p \times 1$ vectors v_t, and compute y_t. The macro data are exact aggregates of the micro data, so that

$$\eta_t = a'y_t \quad \text{and} \quad z_{t-1} = X'_{t-1}a,$$

where the aggregation vector $a = \iota_p/p$. The macro data are thus generated by

$$\eta_t = z'_{t-1}\gamma + u_t, \qquad u_t \sim \mathrm{N}(0, \sigma_u^2), \tag{22}$$

where all k elements of γ are also set equal to one, $u_t = a'v_t$ almost surely, and $\sigma_u^2 = \sigma_v^2 a'a = 1/p$.

4.2. The Model

The base model is the same as the data-generation process, except that the micro and macro statisticians do not know the exact relationship between the two models and hence do not utilize the relationship $u_t = a'v_t$ (almost surely).

The OLS estimator in the micro model (21) is given by

$$\hat{\beta} = \left(\sum X'_{t-1}X_{t-1}\right)^{-1} \sum X'_{t-1}y_t,$$

and the one-period-ahead forecast can be computed as $\hat{y}^{(2)}_{T+1} = X_T\hat{\beta}$. All sums, unless otherwise indicated, are for $t = 1, \ldots, T$. The variance of $\hat{\beta}$ is

$$\Sigma = \mathrm{var}(\hat{\beta}) = \sigma_v^2 X_T \left(\sum X'_{t-1}X_{t-1}\right)^{-1} X'_T.$$

The macro model is given by (22). The OLS estimator for γ is

$$\hat{\gamma} = \left(\sum z_{t-1}z'_{t-1}\right)^{-1} \sum z_{t-1}\eta_t = \left(\sum X'_{t-1}aa'X_{t-1}\right)^{-1} \sum X'_{t-1}aa'y_t,$$

and the one-period-ahead forecast can be computed as $\hat{\eta}^{(1)}_{T+1} = z'_T\hat{\gamma}$. The residuals are defined as

$$\hat{v}_t = y_t - X_{t-1}\hat{\beta}, \qquad \hat{u}_t = \eta_t - z'_{t-1}\hat{\gamma}$$

and the variance estimates as

$$\hat{\sigma}_v^2 = \frac{1}{pT}\sum \hat{v}'_t\hat{v}_t, \qquad \hat{\sigma}_u^2 = \frac{1}{T}\sum \hat{u}_t^2, \qquad \hat{\Sigma} = \hat{\sigma}_v^2 X_T\left(\sum X'_{t-1}X_{t-1}\right)^{-1} X'_T$$

$$\hat{\sigma}_1^2 = \hat{\sigma}_u^2 z'_T\left(\sum z_{t-1}z'_{t-1}\right)^{-1} z_T, \qquad \hat{\sigma}_2^2 = a'\hat{\Sigma}a.$$

One may verify that

$$\pi := \mathrm{cov}(\hat{y}^{(2)}_{T+1}, \hat{\eta}^{(1)}_{T+1}) = \sigma_v^2 X_T\left(\sum X'_{t-1}X_{t-1}\right)^{-1} z_T = \Sigma a,$$

because $\mathrm{cov}(y_t, \eta_t) = \mathrm{cov}(v_t, a'v_t) = \sigma_v^2 a$. Because the micro and macro models are estimated separately, the fact that u_t and v_t are correlated is not known and not used by our macro and micro statisticians. As a result, $\mathrm{cov}(y_t, \eta_t)$ is not estimated by $\hat{\sigma}_v^2 a$ but rather by $\widehat{\text{cov}}(y_t, \eta_t) = (1/T)\sum \hat{v}_t \hat{u}_t$, which is not the same. The estimate for π is then given by

$$\hat{\pi} = X_T\Big(\sum X'_{t-1}X_{t-1}\Big)^{-1}\Big(\sum X'_{t-1}\widehat{\text{cov}}(y_t, \eta_t)z'_{t-1}\Big)\Big(\sum z_{t-1}z'_{t-1}\Big)^{-1}z_T,$$

and, although $\pi = \Sigma a$, it is not true that $\hat{\pi} = \hat{\Sigma} a$. We note that, in practice, we cannot compute this estimate, because our information consists only of the two forecasts and their precisions. In the simulations we can, however, compute the estimate, and it allows us to study the sensitivity of the feasible forecasts. All Monte Carlo experiments were performed with 10,000 replications.

4.3. Forecast Comparisons

We now compare the various forecasts assuming the data-generation process and the models from the base scenario. As a target for comparison, we shall use the forecast (6),

$$\hat{y}_{T+1}(\alpha) = \hat{y}^{(2)}_{T+1} + \frac{\alpha(\hat{\eta}^{(1)}_{T+1} - \hat{\eta}^{(2)}_{T+1})}{a'\hat{\Sigma}a}\hat{\Sigma}a,$$

considered as a function of α. The trace of its mean squared error is

$$\tau^2(\alpha) = \text{tr}(\text{E}(y_{T+1} - \hat{y}_{T+1}(\alpha))(y_{T+1} - \hat{y}_{T+1}(\alpha))'),$$

also a function of α. We define τ_0^2 as the minimum value of $\tau^2(\alpha)$, and we call τ_0 the target root mean squared error (RMSE). We emphasize again that, given the data constraints on our statistician, only three of the six forecasts can actually be computed in practice, namely, the micro forecast and the forecasts whose RMSEs are labeled τ_2 and τ_4. The remaining three forecasts are presented for comparison and sensitivity purposes only.

Table 1a gives the results for the micro forecast and for the five competing forecasts. In each case we present

$$\frac{\tau_i - \tau_0}{\tau_0} \qquad (i = 1, \ldots, 5),$$

that is, the performance of the relevant RMSE relative to the target RMSE. Because $\pi = \Sigma a$ in the base model, we obtain $\hat{y}_{T+1} = \hat{y}^{(2)}_{T+1}$ in both (6) and (12). Hence, the forecast based on micro data only (and ignoring macro information) is the optimal forecast in this case. The relative loss compared to itself (Table 1a(1)) is obviously zero.

We do not present the results for each data set; instead we provide only the minimum, maximum, and median of the relative RMSE among the different data sets. If we replace the optimal α (i.e., α_1) by its estimate $\hat{\alpha}_1$, we obtain τ_1; see Table 1a(2). For most data sets, the results are close to Table 1a(1) because

$\hat{\alpha}_1$ is close to α_1. But for some data sets the relative loss is very large. Setting $\pi = 0$ we obtain $\hat{y}_{T+1}(\alpha_2)$ and τ_2, and the relative loss is reported in Table 1a(3). This should not be as good as Tables 1a(1) and 1a(2) because in fact $\pi \neq 0$, but in fact the performance is better, both in terms of "worst-case scenario" (max) and in terms of the median.

In Tables 1a(4) and 1a(5) we report the RMSE for the forecasts (12) and (15). The forecast (12) does not perform well for some data sets, whereas the forecast (15) does perform well. Table 1a(4) is very close to Table 1a(2), whereas Table 1a(5) is slightly worse than Table 1a(3) because $\alpha_2 < \alpha_4$ because of the bias term.

Finally, the model that incorporates macro information into the micro model leads to forecast (20). Here we estimate parameters that are in fact zero, and hence we lose efficiency. This is reflected in Table 1a(6).

5. DEVIATIONS FROM THE BASE SCENARIO

In real-life applications, the micro model and the macro model will deviate more than in the idealized situation of the previous section. We now investigate four scenarios that mimic situations as they might occur in practice.

5.1. Measurement Errors in the Micro Regressors

We assume that there are two regressors ($k = 2$, data sets 1–5) and that the data are generated as in the base case. However, the second regressor is badly measured at the micro level, so that we observe $x^*_{t-1,2} = x_{t-1,2} + \zeta_{t-1}$ instead of $x_{t-1,2}$. The $\{\zeta_t\}$ are i.i.d. $\mathrm{N}(0, \sigma_\zeta^2)$, and they are independent of all other errors. The estimates and forecast at the micro level are thus based on $X^*_{t-1} := (x_{t-1,1}, x^*_{t-1,2})$. The effect of the measurement error will be that the OLS estimator $\hat{\beta}$ is inconsistent with a bias toward zero.

Measurement errors can either accumulate or dampen when aggregated. Here we assume that they dampen and in fact that measurement errors are absent in the macro model. Thus, the regressors in the macro model are $z_{t-1} = X_{t-1}a$ and not $z_{t-1} = X^*_{t-1}a$. In Figure 2 we plot $\tau(\alpha)$ (solid line) for three values of σ_ζ, namely, 0.0, 0.1, and 0.5. The case $\sigma_\zeta = 0$ is identical to the base scenario discussed in the previous section. Hence, the minimum τ_0 occurs at $\alpha = 0$. The values of $\tau_1, \ldots, \tau_5$ are also plotted, and we see (for this specific data set) that τ_1 performs best and τ_5 worst, in correspondence to Table 1.

If $\sigma_\zeta > 0$, then the minimum of $\tau(\alpha)$ will not be at $\alpha = 0$. In this specific example, τ_4 performs best. This is the simple "Bayesian" solution, ignoring information on π.

Tables 1b and 1c provide more detailed information. When $\sigma_\zeta = 0.1$ (Table 1b), the optimum is not at zero, and the naive micro forecast is not optimal any more. This is reflected in Table 1b(1). Table 1b(2) (reporting on τ_1) is further away from the optimum, because of the misspecification, which

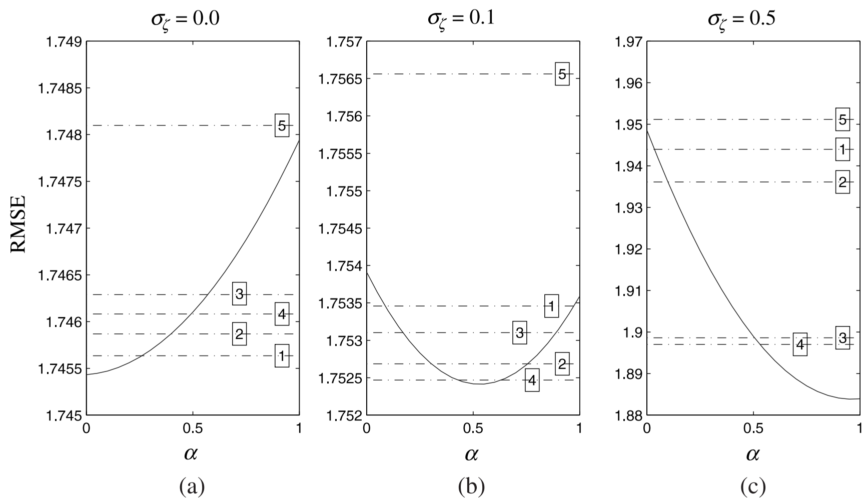

FIGURE 2. Effect of measurement errors in micro regressors, $T = 24$, $p = 3$, data set = 3, for three values of σ_ζ.

implies that α_1 is no longer the optimum value for α. Table 1b(3) is close to the optimum, showing that the wrong assumption $\pi = 0$ is less damaging than the misspecification. Table 1b(4) is again similar to Table 1b(2), but, because the optimal α is not zero, there is no strict dominance in this case. Table 1b(5) (reporting on τ_4) is uniformly closer to the minimum than Table 1b(4), which shows again that information on π is not necessarily useful. The explanation for Table 1b(6) is similar to Table 1a(6) with the addition that we not only estimate zeros in this case but also have a measurement error bias.

In Table 1c we consider the case where $\sigma_\zeta = 0.5$. The differences between the various forecasts are larger now than in the case when $\sigma_\zeta = 0.1$. The naive micro forecast in Table 1c(1) is not a good forecast any more. The correct specification assumption remains more important than the assumption that $\pi = 0$, so that Tables 1c(2) and 1c(4) are dominated by (bigger than) Tables 1c(3) and 1c(5), respectively. The forecast corresponding to τ_4 is the clear winner in this case.[4]

5.2. Nonavailability of Some Micro Regressors

We now assume that there are three regressors ($k = 3$, data sets 6–10). However, the third regressor is only available at the macro level, not at the micro level. As a result we have $X^*_{t-1} := (x_{t-1,1}, x_{t-1,2})$, whereas $X_{t-1} := (x_{t-1,1}, x_{t-1,2}, x_{t-1,3})$. When one of the regressors is not available at the micro level,

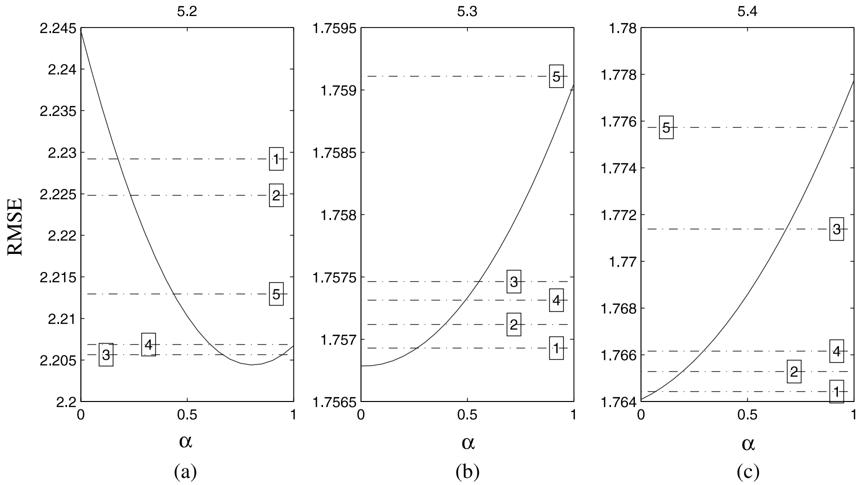

FIGURE 3. Effect of three types of misspecification, $T = 24$, $p = 3$, data set = 8.

then the forecast performance depends on which regressor is not available and how much useful information is lost. Figure 3a illustrates that the minimum of the $\tau(\alpha)$-curve is not at $\alpha = 0$ and gives the RMSEs for the five competing forecasts for one data set. Table 1d provides more detailed information and shows that the loss can be substantial. For example, in Table 1d(1) the naive micro forecast suffers badly from the omitted variable bias, in particular in one of the data sets (number 7). (These outliers are the main reason for using the median rather than the mean in reporting the simulation results.) Tables 1d(2) and 1d(3) are slightly closer to the target, but not as much as Tables 1d(4) and 1d(5), which shows that the bias term in α_1 is poorly estimated. In most cases, Table 1d(4) gives better result than Table 1d(5).

The model that incorporates macro information into the micro model (forecast (20)) is interesting in this case. In the previous scenarios, this forecast was far from the optimum, but here it gives a better result than the minimum of the α-curve. This is not a surprise, because we are using a macro regressor as a proxy for the corresponding micro regressor (which is not available) and this proxy is a very good proxy.

5.3. Nonimportance of Some Macro Regressors

Next we assume that there are three regressors ($k = 3$, data sets 6–10) and that the third regressor satisfies $a'x_{t-1,3} = 0$ for all t. The statistician does not know this, and so the micro model is estimated without imposing the restriction. In the macro model, all observations on $z_{t-1,3}$ will be zero, and so this regressor is dropped from the model and the macro estimates and forecasts are based on two regressors only.

We see from Figure 3b that the minimum of the $\tau(\alpha)$-curve is at $\alpha = 0$. Hence it is no surprise that the forecasts labeled 1 and 2 perform well in this case. We might expect efficiency gains because we estimate only two (rather than three) parameters at the macro level, but the efficiency gain is very small and Figure 3b looks very similar to Figure 2a. Table 1e confirms that the results are similar to the base scenario.

5.4. Correlation at the Micro Level

Finally, it may happen (indeed, it is likely) that the micro units are correlated, so that $v_t \sim \mathrm{N}(0, \Sigma_v)$, where Σ_v is not proportional to the identity matrix. Here we assume that

$$\Sigma_v = (1-\rho)I_p + \rho\iota_p\iota_p' = \theta_1 J + \theta_2(I_p - J),$$

where $J := \iota_p\iota_p'/p$ is an idempotent matrix of rank one and the coefficients are $\theta_1 := 1 + (p-1)\rho$ and $\theta_2 := 1 - \rho$. The statistician, however, is unaware of the correlation. We take $\rho = 0.5$. The data-generation process has changed, but the models are still the same.

In this case the theoretical minimum is often obtained for $\alpha < 0$; see Figure 3c. We know that this is theoretically possible, and truncation at zero would lead to poorer performance. The forecasts reported as τ_1 and τ_3 perform poorly; the other forecasts perform reasonably well.

6. EXTENSION TO TIME SERIES

Let us summarize our findings so far. Our purpose was to construct an improved micro forecast $\hat{y}_{T+1}$ in the situation where we are given a macro forecast $\hat{\eta}^{(1)}_{T+1}$ with estimated variance $\hat{\sigma}_1^2$ and a micro forecast $\hat{y}^{(2)}_{T+1}$ with estimated variance matrix $\hat{\Sigma}$ and where, in addition, we wish the improved forecasts to satisfy a constraint $a'\hat{y}_{T+1} = \hat{\eta}_{T+1}$. Based on the preceding theory and simulations we propose the micro forecast

$$\hat{y}_{T+1} = \hat{y}^{(2)}_{T+1} + \frac{\alpha(\hat{\eta}^{(1)}_{T+1} - a'\hat{y}^{(2)}_{T+1})}{\hat{\sigma}_2^2}\hat{\Sigma}a, \tag{23}$$

where $\hat{\sigma}_2^2 = a'\hat{\Sigma}a$ and $\alpha = \hat{\sigma}_2^2/(\hat{\sigma}_1^2 + \hat{\sigma}_2^2)$ denotes the weight in the associated macro forecast

$$\hat{\eta}_{T+1} = \alpha\hat{\eta}^{(1)}_{T+1} + (1-\alpha)a'\hat{y}^{(2)}_{T+1}. \tag{24}$$

The proposed micro forecast has a minimum mean squared error interpretation and is also optimal in a least squares and a Bayesian framework.

Table 1. Relative RMSE comparisons (in %) for various forecasts and various scenarios

		(1) Micro forecast			(2) τ_1			(3) τ_2			(4) τ_3			(5) τ_4			(6) τ_5		
T	p	min	max	med	min	max	med	min	max	med	min	max	med	min	max	med	min	max	med
(a) Base scenario																			
24	3	0.00	0.00	0.00	0.00	364.00	0.01	0.00	0.14	0.03	0.00	321.90	0.08	0.00	0.29	0.09	0.15	1.45	0.66
24	9	0.00	0.00	0.00	0.00	38.90	0.14	0.00	0.01	0.00	0.00	7.38	0.15	0.00	0.04	0.00	0.03	0.28	0.12
96	3	0.00	0.00	0.00	0.00	12.60	0.00	0.00	0.00	0.00	0.00	17.86	0.00	0.00	0.02	0.00	0.01	0.17	0.02
96	9	0.00	0.00	0.00	0.00	0.77	0.00	0.00	0.01	0.00	0.00	160.36	0.00	0.00	0.02	0.01	0.01	0.17	0.06
(b) Measurement error in micro regressors, $\sigma_\zeta = 0.1$																			
24	3	0.01	0.17	0.01	0.01	224.33	0.06	0.01	0.07	0.02	0.02	141.26	0.57	0.00	0.19	0.04	0.24	1.44	0.82
24	9	0.00	0.06	0.05	0.00	840.34	21.84	0.00	0.02	0.01	0.01	109.50	3.38	0.00	0.03	0.01	0.09	0.29	0.14
96	3	0.01	0.17	0.17	0.01	327.39	3.79	0.01	0.10	0.07	0.05	414.94	11.74	0.03	0.04	0.04	0.09	0.34	0.18
96	9	0.00	0.06	0.01	0.00	5.84	0.01	0.00	0.03	0.00	0.00	168.30	0.02	0.00	0.02	0.00	0.06	0.19	0.12
(c) Measurement error in micro regressors, $\sigma_\zeta = 0.5$																			
24	3	1.14	3.87	2.75	1.00	32.31	2.36	0.85	2.77	1.82	0.48	298.75	2.82	0.12	0.71	0.52	0.97	3.81	3.44
24	9	0.07	1.31	1.08	0.07	74.66	3.04	0.04	0.88	0.80	0.14	465.47	121.73	−0.03	0.24	0.20	0.28	1.20	1.15
96	3	0.10	3.66	3.50	0.08	5.03	3.50	0.08	3.32	2.99	2.03	311.07	15.39	0.66	0.78	0.69	0.04	3.59	3.20
96	9	0.55	1.21	1.09	0.55	1.15	1.04	0.52	1.09	0.92	0.49	18.21	0.73	0.24	0.41	0.32	0.42	1.18	0.98
(d) Nonavailability of micro regressors																			
24	3	0.07	197.95	1.82	−0.01	147.22	1.13	−0.04	147.15	0.93	0.00	1.39	0.33	0.01	1.74	0.16	−29.12	29.47	0.39
24	9	0.19	2,368.86	4.42	0.02	1,832.12	4.20	0.02	1,832.06	4.06	0.00	1.50	0.78	0.00	2.00	1.03	−3.84	352.05	0.00
96	3	0.68	674.77	2.38	0.03	143.78	1.85	0.05	143.80	1.60	0.00	6.82	0.01	0.00	7.85	0.24	0.00	659.38	1.45
96	9	3.31	3,837.21	7.61	3.25	2,768.53	7.56	3.20	2,768.53	7.53	0.00	6.09	0.67	0.00	5.41	1.02	−4.50	23.11	0.60
(e) Macro restriction																			
24	3	0.00	0.00	0.00	0.00	43.92	0.05	0.00	0.11	0.02	0.02	31.11	0.74	0.00	0.23	0.03	0.05	1.33	0.60
24	9	0.00	0.00	0.00	0.00	1.88	0.02	0.00	0.01	0.00	0.01	8.61	0.12	0.00	0.03	0.00	0.00	0.25	0.12
96	3	0.00	0.00	0.00	0.00	1.55	0.00	0.00	0.00	0.00	0.00	88.44	0.00	0.00	0.00	0.00	0.00	0.18	0.02
96	9	0.00	0.00	0.00	0.00	0.64	0.00	0.00	0.01	0.00	0.00	7.03	0.00	0.00	0.02	0.00	0.00	0.16	0.06
(f) Micro correlation																			
24	3	0.02	0.19	0.17	0.03	63,591.07	518.56	0.02	0.46	0.25	0.02	10,293.19	5.33	0.02	0.66	0.39	0.33	3.32	0.94
24	9	0.00	0.15	0.01	0.02	3,200.58	0.17	0.00	0.18	0.04	0.00	431.96	0.09	0.00	0.23	0.05	−0.07	2.19	0.74
96	3	0.00	0.03	0.00	0.00	653.69	0.01	0.00	0.05	0.01	0.00	0.32	0.00	0.00	0.07	0.01	−0.01	0.09	0.06
96	9	0.00	0.03	0.01	0.00	1,288.70	0.12	0.00	0.03	0.02	0.00	114.77	0.12	0.00	0.05	0.03	0.03	1.21	0.30

In a time-series framework the aggregate level (e.g., a quarterly model) forecasts the average outcome for several periods (say, the average over three months in one quarter), whereas the more detailed level (a monthly model) forecasts each period (month) separately. In this sense one can think about the aggregated (quarterly) level as a "macro" model and about the detailed (monthly) level as a "micro" model. This provides the link between the current section and the previous sections.

Implicit in the theory developed so far is that the data become available *once* per period, so that the additional data required to forecast period $T+1$ all become available at the same time, namely, at the end of period T. This is reasonable if we consider p micro units in one macro unit, such as p firms in one industry or p industries in one sector or p sectors in one economy. But it is not reasonable in a time-series framework where we consider $p = 3$ months in one quarter or $p = 4$ quarters in one year. In that case the relevant micro data will become available during the year and not just at the end of the year. An extension of the theory is thus required.

The extension is straightforward. Suppose that p_1 micro periods have passed so that $p - p_1$ micro periods are to be forecasted. The first p_1 elements of y_{T+1} are then known so that

$$\hat{y}_{T+1}^{(2)} = (y_{T+1,1}, \ldots, y_{T+1,p_1}, \hat{y}_{T+1,p_1+1}^{(2)}, \ldots, \hat{y}_{T+1,p}^{(2)})'.$$

The estimated variance matrix of this vector has the form

$$\hat{\Sigma} = \begin{pmatrix} 0 & 0 \\ 0 & \hat{\Sigma}_2 \end{pmatrix},$$

where Σ_2 is a $(p - p_1) \times (p - p_1)$ matrix. At the macro level there is no new information, and so there is no need to redefine the original macro forecast $\hat{\eta}_{T+1}^{(1)}$.

With these modifications of $\hat{y}_{T+1}^{(2)}$ and $\hat{\Sigma}$ we can compute the micro forecast $\hat{y}_{T+1}$ and the macro forecast $\hat{\eta}_{T+1}$ from (23) and (24), and the same optimality results will hold.

As we proceed through period $T+1$, the original macro forecast $\hat{\eta}_{T+1}^{(1)}$ does not change, but the vector of micro forecasts $\hat{y}_{T+1}^{(2)}$ is updated together with its variance $\hat{\Sigma}$. This will affect α. One would expect that the weight of the micro information increases when more information becomes available, so that α decreases. To demonstrate this intuition assume that $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2)$ is diagonal and known and that $a = \iota_p/p$ where ι_p denotes the $p \times 1$ vector of ones. Then α as a function of p_1 will be

$$\alpha(p_1) = \frac{a'\hat{\Sigma}a}{\hat{\sigma}_1^2 + a'\hat{\Sigma}a} = \frac{\sigma_{p_1+1}^2 + \cdots + \sigma_p^2}{p^2\hat{\sigma}_1^2 + \sigma_{p_1+1}^2 + \cdots + \sigma_p^2},$$

which obviously gives $\alpha(p_1) > \alpha(p_1 + 1)$.

7. IMPROVING THE EURO YIELD FORECAST BY COMBINING MONTHLY AND QUARTERLY MODELS

We demonstrate our theory by presenting new predictions of European interest rates, in particular zero rates; see Figure 4. As predictors we take various macro and financial series. A detailed description of the data set is provided in Appendix B.

An important feature of the European series is its relative shortness. The series starts with the introduction of the Euro in January 1999, which gives us seven full years of observations. In addition, some of the macro series are released with a half-year delay, so that the final data set consists of 78 monthly observations and 26 quarterly observations. This is an extremely short series when compared to the U.S. series, which is available from the early 1960s. There is some evidence of convergence of the yield curves of the Euro-zone countries (see, e.g., Batten and Fetherston, 2004), but this does not allow us to extend the European series significantly.

It is therefore important to use as much information as possible to get a good forecast. The annual observations cannot provide a reliable forecast, but there are enough quarterly observations to construct a reasonable model, especially because they are much less noisy than the monthly data.

The shortness of the series also explains why relatively few papers exist on the European market. In contrast, the U.S. market has been studied extensively; see Barrett, Gosnell, and Heuson (2004), Piazzesi and Swanson (2004), and Diebold, Rudebusch, and Aruoba (2006). An excellent survey and empirical comparisons can be found in Diebold and Li (2006).

We shall be concerned with the prediction of maturities of 1, 2, and 5 years, rather than with the term structure of interest rates. The main reason is that bonds with higher maturities are not exercised during the observed period, so that there is no evidence of how well they were priced. We predict the selected yields directly rather than using factor models, because each additional step

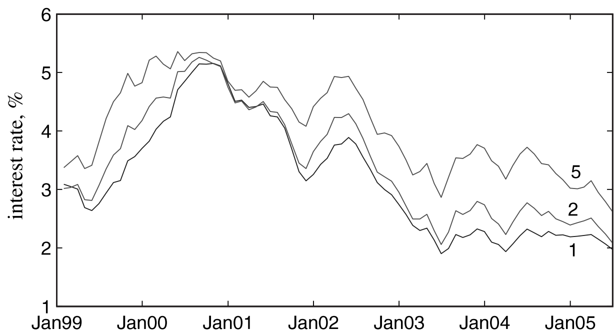

FIGURE 4. Euro zero rates for maturities 1, 2, and 5 years.

TABLE 2. RMSE of separate forecasts

Maturity	Monthly forecast	Quarterly forecast
1	0.175	0.602
2	0.204	0.582
5	0.205	0.454

(extracting factors in this case) consumes information and the information is very limited in our case because of the short history. Model selection and pretesting issues are ignored. The final models are similar to the models typically used for the U.S. yields. The final improvement of the forecast is uniform over different possible sets of monthly and quarterly models.

To compare different forecasts we look again at the RMSE. The first half of the sample is used for estimation, and we use the second half for out-of-sample forecasting. The forecasts are based on the linear models (1) and (2) with the regressors described in Appendix B. For both models we use a one-step-ahead linear forecast with an increasing window. The results for the separate monthly and quarterly forecasts are given in Table 2. One should not conclude from Table 2 that the monthly model is better than the quarterly model, because they are not nested. Moreover, they are structurally different and use a different number of observations and predictors. However, one can see that the monthly model forecasts short-term yields better than long-term yields, whereas the quarterly model is more accurate for the long-term yields. Notice that in this case micro and macro data are both available but that the micro data do not become available at once—they come in three installments.

The quarterly yield is just the average of the monthly yields, but for the forecasts this is not necessarily the case. The procedure from Section 6 allows us to bring the quarterly and monthly forecasts in accordance with each other by taking the restriction vector to be $a = (\frac{1}{3})(1, 1, 1)'$. Combining monthly and quarterly forecasts we obtain the results presented in Table 3.

The updating procedure improves the monthly forecast. One can argue that the improvement is not significant, but if it is translated into basis points (bp: 100 bp = 1%), which is a standard measure for the bond market, then the improvement is 40, 100, and 180 bp, respectively, for 1, 2, and 5 years to maturity, which is considerable.

This is for the monthly ("micro") forecasts. For the quarterly ("macro") forecasts we have three updated versions; see Table 4. The first update (I) is made at the beginning of the quarter, when the monthly forecasts are made with one-, two-, and three-step-ahead models. At the beginning of the second month, the first month has been observed, and one- and two-step-ahead forecasts are used

Table 3. RMSE comparison of the monthly forecasts

Maturity	Updated monthly forecast	Monthly forecast
1	0.171	0.175
2	0.194	0.204
5	0.187	0.205

for the second and third months. This gives the second update (II). Finally, the third update (III) is made at the beginning of the third month when the values for the first and second months are known; the third month is forecasted with a one-step-ahead model. As Table 4 reveals, with the new monthly information received we can improve the quarterly forecast considerably, and the more information revealed, the more accurate our forecast becomes. This simple illustration thus demonstrates the importance of combining information at different levels and also suggests that our proposed method works well in practice.

8. SUMMARY AND CONCLUSIONS

The purpose of this paper was to construct an improved micro forecast $\hat{y}_{T+1}$ in the situation where we are given a macro forecast $\hat{\eta}^{(1)}_{T+1}$ with estimated variance $\hat{\sigma}_1^2$ and a micro forecast $\hat{y}^{(2)}_{T+1}$ with estimated variance matrix $\hat{\Sigma}$ and where, in addition, we wish the improved forecasts to satisfy a constraint $a'\hat{y}_{T+1} = \hat{\eta}_{T+1}$. Based on the theory and simulations in this paper we propose the micro forecast (23) and the macro forecast (24). We emphasize again that only the *relative* precisions of the forecasts are required.

The proposed forecasts are simple and robust. The robustness is investigated in the simulations in two ways. First, by assuming (τ_1 and τ_3) that we can estimate the correlation π, which in most cases we can not. The additional information does not appear to help us; in fact, the contrary is the case. Second, by

Table 4. RMSE comparison of the quarterly forecasts

Maturity	Updated quarterly forecast, I	Updated quarterly forecast, II	Updated quarterly forecast, III	Quarterly forecast
1	0.266	0.151	0.063	0.602
2	0.309	0.183	0.068	0.582
5	0.288	0.192	0.062	0.454

assuming that *all* data are available and that we incorporate the macro regressors in the micro model (τ_5); this too does not appear to be a fruitful extension.

The proposed forecast performs uniformly well, both on average (considering the median relative loss of the various data sets) and from the worst-case scenario (maximum relative loss) viewpoint. Various types of misspecification were considered: micro measurement error, absence of one of the micro regressors, a macro restriction, and micro correlation. In all these cases our proposed forecast performs well.

An empirical example concerning forecasts of the Euro yield curve confirms that a time-series modification of our procedure improves both monthly and quarterly forecasts.

The method can be generalized to more than two levels, to nonlinear models such as factor models (Stock and Watson, 2002a, 2002b), to nonlinear restrictions, and to nonnormal densities.

NOTES

1. The study of perfect aggregation can be traced back to Theil (1954). This aggregation is very restrictive, and "perfect aggregation" tests usually favor the disaggregated model; see Pesaran, Pierse, and Kumar (1989).

2. There are several reasons why the macro model may not be just a (weighted) average (or sum) of the micro models. First, the macro observations are usually measured more precisely than the micro observations. Second, the transition from one level to another can be nonlinear. For example, some shocks are dampened in the long run.

3. Instead of the trace we might also consider minimizing $\mathrm{MSE}(a'\hat{y}_{T+1})$. This leads to $\alpha_1^* = a'c/a'q$. One may verify that $\alpha_1 = \alpha_1^*$ if and only if $(a'\Sigma a)(a'\Sigma\pi) = (a'\Sigma^2 a)(a'\pi)$, which is the case, for example, when $\pi = 0$ or $\Sigma = I$ or $\Sigma a = \pi$.

4. Notice the negative number in Table 1c(5), implying that $\tau_4 < \tau_0$ in one of the data sets (number 4). This can happen, because the α-curve and therefore the theoretical optimum is calculated by averaging the simulation results, keeping α fixed. When α is estimated, then the average is not necessarily on the α-curve (this is the reason why on the pictures, when α is estimated, the results are shown as a straight line rather than a point). Therefore it is possible that $\tau_1, \ldots, \tau_4$ are less than the simulated optimum. This is, however, a rare event.

REFERENCES

Barrett, W.B., T.F. Gosnell, & A.J. Heuson (2004) Term-structure factor shifts and economic news. *Financial Analysts Journal* 60, 81–94.

Bates, J.M. & C.W.J. Granger (1969) The combination of forecasts. *Operational Research Quarterly* 20, 451–468.

Batten, J.A. & T.A. Fetherston (2004) *European Fixed Income Markets: Money, Bond, and Interest Rate Derivatives*. Wiley.

Clemen, R.T. (1989) Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5, 559–583.

Clements, M.P. & D.F. Hendry (1993) On the limitations of comparing mean square forecast errors. *Journal of Forecasting* 12, 617–637.

Diebold, F.X. & L. Kilian (2001) Measuring predictability: Theory and macroeconomic applications. *Journal of Applied Econometrics* 16, 657–669.

Diebold, F.X. & C. Li (2006) Forecasting the term structure of government bond yields. *Journal of Econometrics* 130, 337–364.
Diebold, F.X. & J.A. Lopez (1996) Forecast evaluation and combination. In G.S. Maddala & C.R. Rao (eds.), *Handbook of Statistics*, pp. 214–268. North-Holland.
Diebold, F.X., G.D. Rudebusch, & S.B. Aruoba (2006) The macroeconomy and the yield curve: A dynamic latent factor approach. *Journal of Econometrics* 131, 309–338.
Hendry, D.F. & M.P. Clements (2004) Pooling of forecasts. *Econometrics Journal* 7, 1–31.
Hendry, D.F. & K. Hubrich (2005) Forecasting Aggregates by Disaggregates. Discussion paper, European Central Bank.
Hoeting, J.A., D. Madigan, A.E. Raftery, & C.T. Volinsky (1999) Bayesian model averaging: A tutorial. *Statistical Science* 14, 382–401.
Imbens, G.W. & T. Lancaster (1994) Combining micro and macro data in microeconometric models. *Review of Economic Studies* 61, 655–680.
Inoue, A. & L. Kilian (2006) On the selection of forecasting models. *Journal of Econometrics* 130, 273–306.
Lancaster, T. (2004) *An Introduction to Modern Bayesian Econometrics*. Blackwell.
Leung, G. & A.R. Barron (2006) Information theory and mixing least-squares regressions. *IEEE Transactions on Information Theory* 52, 3396–3410.
Magnus, J.R., J.W. van Tongeren, & A.F. de Vos (2000) National accounts estimation using indicator ratios. *Review of Income and Wealth* 46, 329–350.
Marcellino, M., J.H. Stock, & M.W. Watson (2003) Macroeconomic forecasting in the Euro area: Country specific versus area-wide information. *European Economic Review* 47, 1–18.
Pesaran, M.H., R.G. Pierse, & M.S. Kumar (1989) Econometric analysis of aggregation in the context of linear prediction models. *Econometrica* 57, 861–888.
Piazzesi, M. & E. Swanson (2004) Futures Prices as Risk-Adjusted Forecast of Monetary Policy. NBER Working paper 10547.
Raftery, A.E., D. Madigan, & J.A. Hoeting (1997) Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92, 179–191.
Stock, J.H. & M.W. Watson (2002a) Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics* 20, 147–162.
Stock, J.H. & M.W. Watson (2002b) Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167–1179.
Theil, H. (1954) *Linear Aggregation of Economic Relations*. North-Holland.
Yang, Y. (2004) Combining forecasting procedures: Some theoretical results. *Econometric Theory* 20, 176–222.

APPENDIX A: Estimation of the Common Mean of a Bivariate Distribution with Known Variance

Suppose we have two measurements on an unknown quantity θ:

$$y_1 = \theta + \varepsilon_1, \quad y_2 = \theta + \varepsilon_2,$$

where $(\varepsilon_1, \varepsilon_2)$ follows a bivariate normal distribution with mean zero and known variance

$$\operatorname{var}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

If ε_1 and ε_2 are uncorrelated ($\sigma_{12} = 0$) and the two variances are equal, then we estimate θ by $\hat{\theta} = (y_1 + y_2)/2$. If ε_1 and ε_2 are uncorrelated and the two variances are not equal, then we estimate θ by

$$\hat{\theta} = \alpha y_1 + (1 - \alpha) y_2, \qquad \alpha = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2},$$

which is a weighted average of y_1 and y_2, where the weight α lies between zero and one.

Now consider the case where ε_1 and ε_2 are correlated. Then we also obtain $\hat{\theta} = \alpha y_1 + (1 - \alpha) y_2$ but now with

$$\alpha = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}.$$

We see that α does *not* necessarily lie between zero and one. In particular, $\alpha < 0$ if and only if $\sigma_{12} > \sigma_2^2$, and $\alpha > 1$ if and only if $\sigma_{12} > \sigma_1^2$. Figure A.1 illustrates this situation. It presents one constant-probability contour for the case $\sigma_1 = 1$, $\sigma_2 = 2$, and $\sigma_{12} = 1.5$. Hence the correlation is $r = \frac{3}{4}$. Our two observations take the values $y_1 = 2$ and $y_2 = 4$, so that we obtain $\alpha = 1.25$ and $\hat{\theta} = 1.5$, which clearly lies outside the interval (y_1, y_2).

At first glance this may seem unsatisfactory. We have two observations 2 and 4 and an estimate of 1.5. At second glance, however, it becomes clear that this is the correct solution and that we should not force the solution to be between the two observations.

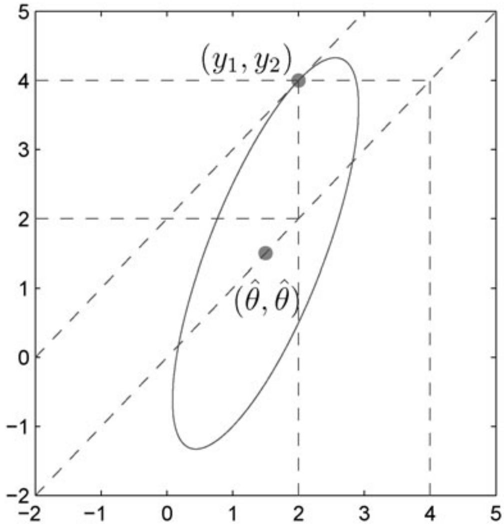

FIGURE A.1. Constant-probability contour for the case $\text{cov}(y_1, y_2) > \text{var}(y_1)$.

Consider the extreme situation where $r = 1$ and $\sigma_1 \neq \sigma_2$. Then we have $y_1 = \theta + \sigma_1 \varepsilon$ and $y_2 = \theta + \sigma_2 \varepsilon$, where the common noise ε satisfies $\varepsilon \sim \mathrm{N}(0,1)$. In this case, $\hat{\theta}$ *must* lie outside the interval (y_1, y_2). We simply solve the two equations in two unknowns (θ and ε) and find $\hat{\theta} = \alpha y_1 + (1-\alpha) y_2$ with

$$\alpha = \frac{\sigma_2}{\sigma_2 - \sigma_1}.$$

The "weight" α in this case is either larger than one (if $\sigma_1 < \sigma_2$) or smaller than zero (if $\sigma_1 > \sigma_2$).

Consider another situation where

$$y_1 \sim \mathrm{N}(\theta, \sigma_1^2), \qquad y_2 = y_1 + \varepsilon,$$

where ε has mean zero and is distributed independently of y_1. In this case $\mathrm{cov}(y_1, y_2) = \sigma_1^2$ and $\hat{\theta} = y_1$. The observation y_1 is a sufficient statistic for θ, and the information contained in y_2 is superfluous.

We conclude that—in the presence of correlation—"weights" may lie outside the $(0,1)$ interval.

APPENDIX B: The Data

The data used in the empirical example of Section 7 have been obtained from Eurostat Web site http://europa.eu.int/comm/eurostat and cover the period from January 1999 until June 2005. We analyze the 12 Euro-zone countries Belgium, Germany, Greece (from 1 January 2001), Spain, France, Ireland, Italy, Luxembourg, the Netherlands, Austria, Portugal, and Finland.

Some of the series, especially indices, contain a unit root. The formal unit root analysis is presented in Table B.1. Following Piazzesi and Swanson (2004) we use year-on-year log differences to remove the unit root. This is convenient, because the transformation

TABLE B.1. Results of Augmented Dickey–Fuller (ADF) test

	ADF statistic of original series (and its p-value)	ADF statistic of modified series (and its p-value)
HICP	0.22 (0.97)	−3.11 (0.03)
IDOP	−1.40 (0.58)	−2.33 (0.17)
GWS	−1.20 (0.67)	−1.97 (0.30)
RTT	−2.63 (0.09)	−0.98 (0.76)
DUR	−0.48 (0.89)	−1.55 (0.50)
IND	−2.63 (0.09)	−1.58 (0.49)
GDP	−1.31 (0.61)	−1.62 (0.45)
CONSTR	−4.75 (0.00)	−1.77 (0.38)

is uniform for monthly and quarterly data. Moreover, the transformation removes possible traces of any seasonal component.

The monthly model includes the 15 predictors described subsequently and a constant.

HICP: Inflation is captured by the consumer price indices, measured for each country separately and combined in one harmonized index. We use the harmonized index provided by Eurostat. The series contains a unit root, and so we use its year-on-year log change.

IDOP: Producer prices are measured by the index of industrial domestic output prices. The series contains a unit root, and so we use its year-on-year log change.

GWS: As an analogue of the nonfarm payrolls in the United States, we use the seasonally adjusted index of gross wages and salaries of total industry excluding construction. The series is again nonstationary, and we use its year-on-year log change.

HUNE: Unemployment is measured by the seasonally adjusted harmonized unemployment index of all age classes including males and females. Similar to the price index it is measured for each country separately and then combined into one harmonized measure using the weighted sum transformation.

RTT: The condition of industry and services is reflected by the seasonally adjusted retail trade turnover index. The series contains a unit root, and so we use its year-on-year log change.

DUR: The seasonally adjusted index of the industrial production of consumer durables is differentiated to remove the unit root.

IND: Stationary adjusted version of the industrial production index for the total industry excluding construction.

DR: The relation between the official deposit rate (DR), the official refinancing operation rate, and the official lending rate is kept fixed (with the exception of a short period January–March 1999) by the Central Bank with a gap of 1%. We arbitrarily choose the official deposit rate for our analysis.

EURIBOR, LIBOR: To account for other possibilities for investment we include money market short-term interest rates EURIBOR for Euro contracts and LIBOR for interbank loans in London. The LIBOR rate is taken because of the high influence this market has on the whole of Europe.

GBP, USD: We take account of international competition for investments by including the exchange rate for pound sterling (GBP) and the U.S. dollar (USD).

SPRD (three predictors): Piazzesi and Swanson (2004) find the yield spreads particularly useful for their analysis. We include three yield spreads: the spread between 2 and 1 year yields ($SPRD_{2,1}$), between 5 and 2 years ($SPRD_{5,2}$), and between 10 and 5 years ($SPRD_{10,5}$).

The quarterly model resembles the yields-macro model of Diebold et al. (2006), which includes real activity, monetary policy, and inflation. It contains a constant, the average inflation, and the average deposit rate from the monthly data and the following three predictors:

GDP: The stationary transformation of the seasonally adjusted gross domestic product in current prices.

CONSTR: The stationary transformation of the seasonally adjusted construction production index.

CU: Percentage of the current level of capacity utilization (CU) from business surveys.