Adjusting Band-Regression Estimators for Prediction: Shrinkage and Downweighting

Erhard Reschenhofer[*], Marek Chudy

Department of Statistics and Operations Research, University of Vienna, Vienna, Austria
*Corresponding author: erhard.reschenhofer@univie.ac.at

Abstract   This paper proposes further developments of band-regression models for forecasting purposes, namely a simple method for shrinking the parameter estimates as well as a method for the automatic selection of the underlying frequency band. In combination with a method for downweighting older data, the improved band-regression model is used to forecast real GDP growth across nine industrialized economies. The results of this empirical study show that this forecasting approach outperforms conventional forecasting methods. As a secondary finding, the empirical results also raise doubts whether the yield-curve spread is really a valuable leading indicator of GDP growth.

Keywords: forecasting, optimal weights, band regression, shrinkage, fractional degrees of freedom

Cite This Article: Erhard Reschenhofer, and Marek Chudy, "Adjusting Band-Regression Estimators for Prediction: Shrinkage and Downweighting." *International Journal of Econometrics and Financial Management*, vol. 3, no. 3 (2015): 121-130. doi: 10.12691/ijefm-3-3-3.

1. Introduction

A robust and often successful forecasting strategy is to combine forecasts from different models, either by averaging [1] or by weighted averaging [2]. Combining forecasts can also be beneficial in situations where there is only one model, but there is uncertainty about the optimal time period used for forecasting [3,4]. In the absence of structural breaks, the use of the full sample period is clearly the method of choice. But if there are breaks, the forecast based on the full sample will be biased. In this case, one could try to estimate the locations of the breaks [5,6,7,8] and use only data from the most recent regime for prediction. Unfortunately, it is often very difficult to obtain precise estimates of the locations and the sizes of the breaks [9]. The problem intensifies when predictions have to be made shortly after some parameter has changed. In general, the forecast based on a recent subsample will have a smaller bias and a higher variance than that based on the full sample. Trying to balance the trade-off between bias and variance, Pesaran and Timmermann [3] proposed to combine forecasts based on subsamples of different lengths. In the context of forecasting random walks with breaks in the drift, Pesaran and Pick [10] obtained Monte Carlo results as well as empirical results indicating that averaging forecasts over estimation windows can improve the forecasting performance compared to a single estimation window when the break size is not too small compared to the volatility. However, Steinberger [11] showed that, depending on the specification of the volatility process, single-window forecasting is asymptotically as good as average-window forecasting or

even better. Pesaran et al. [12] derived "robust optimal" weights, which do not require knowledge of the break dates, and applied them to forecast real GDP growth across nine industrialized economies using the yield-curve spread as a predictor. In this application, their weights outperformed other weights in different settings.

The GDP forecasts in [12] are obtained in two steps. First, the observations are weighted and then conventional regression models are applied to the weighted observations. The main focus of our paper is on the second step. We modify conventional regression models by imposing frequency-domain restrictions which are plausible for economic data and are also supported by empirical evidence. Our approach is related to the band-regression approach which is based on the assumption that linear relationships between several variables exist only in certain frequency bands [13,14]. For example, band regression is used for the analysis of relationships in the presence of seasonal patterns or for the estimation of cointegrating relationships (e.g. [15]). In the former case, narrow bands around all seasonal frequencies must be excluded and, in the latter case, the focus should be on a narrow band close to frequency 0. However, band regression can - in contrast to our approach - not be used for prediction in the time domain.

Section 2.1 derives a shrinkage version of the conventional least squares (LS) estimator under certain frequency-domain restrictions. In addition, fractional degrees of freedom are defined, which are later used in Section 3 for the automatic choice of the frequency-domain restrictions. Section 2.2 discusses different weights which are intended to take care of parameter instability. Section 3 combines the methods discussed in 2.1 and 2.2 and applies them to the economic data set used

in [12]. All methods required for this empirical analysis were programmed in R [16]. Section 4 concludes.

2. Methods

2.1. Band Regression

2.1.1. Deterministic Regressors

Let

$$y = X\beta + u, \tag{1}$$

where $Eu = 0$, $Var(u) = \sigma^2 I$, and X is a nonstochastic $n \times k$ matrix with full column rank $k < n$. Pre-multiplication of both sides of this equation by the matrix H_r', $r > k/2$, which consists of the first $2r$ rows of the matrix

$$H' = \sqrt{\frac{2}{n}} \begin{pmatrix} \cos(\omega_1 1) & \cdots & \cos(\omega_1 n) \\ \sin(\omega_1 1) & \cdots & \sin(\omega_1 n) \\ \vdots & & \vdots \\ \cos(\omega_m 1) & \cdots & \cos(\omega_m n) \\ \sin(\omega_m 1) & \cdots & \sin(\omega_m n) \end{pmatrix}, \tag{2}$$

where $\omega_j = 2\pi j/n$, $j = 1,...,m = [(n-1)/2]$, gives

$$\underbrace{H_r' y}_{y_r} = \underbrace{H_r' X}_{X_r} \beta + \underbrace{H_r' u}_{u_r} \tag{3}$$

with $Eu_r = 0$ and

$$Var(u_r) = EH_r' uu' H_r = \sigma^2 H_r' H_r = \sigma^2 I. \tag{4}$$

We assume that all variables are mean-corrected, hence there is no need to include a constant dummy variable which would produce a column of zeroes in X_r. Alternatively, this singularity could be avoided by including frequency zero, i.e., adding the row

$$\left(\sqrt{\frac{1}{n}} \quad \cdots \quad \sqrt{\frac{1}{n}} \right). \tag{5}$$

In case of odd n, the matrix H would then have n columns. In case of even n, this can also be achieved by including additionally frequency π, i.e., adding also the column

$$\left(-\sqrt{\frac{1}{n}} \quad \sqrt{\frac{1}{n}} \quad -\sqrt{\frac{1}{n}} \quad \cdots \quad \sqrt{\frac{1}{n}} \right). \tag{6}$$

While the estimator

$$\hat{\beta}_r = (X_r' X_r)^{-1} X_r' y_r \tag{7}$$

is best linear unbiased, it will be inefficient relative to

$$\hat{\beta} = (X'X)^{-1} X'y \tag{8}$$

if the rank of H_r' is less than n. Hence, the latter estimator should always be preferred to the former unless it is suspected that some model assumptions are violated.

Replacing both the dependent variable and all regressors by their projections onto the span of the columns of H_r also leads to the band-regression estimator, i.e.,

$$((P_r X)' P_r X)^{-1} (P_r X)' P_r y$$
$$= \left((H_r H_r' X)' H_r H_r' X \right)^{-1} (H_r H_r' X)' H_r H_r' y \tag{9}$$
$$= \left(X' H_r H_r' X \right)^{-1} X' H_r H_r' y = \hat{\beta}_r,$$

where

$$P_r = H_r (H_r' H_r)^{-1} H_r' = H_r H_r'. \tag{10}$$

The covariance matrix of $P_r u$ is given by

$$Var(P_r u) = EH_r H_r' uu' H_r H_r' = \sigma^2 P_r. \tag{11}$$

Now assume that the linear regression model (1) is misspecified because a linear relationship exists only between $\underline{y}_r = P_r y$ and $\underline{X}_r = P_r X$, but not between $\overline{y}_r = (I - P_r)y$ and $\overline{X}_r = (I - P_r)X$, i.e.,

$$E\underline{y}_r = \underline{X}_r \beta_r \tag{12}$$

for some $\beta_r \in \Re^k$, whereas

$$E\overline{y}_r = 0. \tag{13}$$

Under (13), it seems natural to modify the conventional LS estimator

$$\hat{\beta} = (X'X)^{-1} X'y$$
$$= \left((\underline{X}_r + \overline{X}_r)'(\underline{X}_r + \overline{X}_r) \right)^{-1} \left((\underline{X}_r + \overline{X}_r)'(\underline{y}_r + \overline{y}_r) \right) \tag{14}$$
$$= \left(\underline{X}_r' \underline{X}_r + \overline{X}_r' \overline{X}_r \right)^{-1} \left(\underline{X}_r' \underline{y}_r + \overline{X}_r' \overline{y}_r \right)$$

by setting

$$\overline{X}_r' \overline{y}_r = 0. \tag{15}$$

The resulting estimator

$$\tilde{\beta}_r = \left(\underline{X}_r' \underline{X}_r + \overline{X}_r' \overline{X}_r \right)^{-1} \underline{X}_r' \underline{y}_r \tag{16}$$

is a compromise between the band-regression estimator $\hat{\beta}_r$, which is suitable for $\underline{y}_r$, and the trivial estimator $\beta_0 = 0$, which is suitable for $\overline{y}_r$.

2.1.2. Jointly Stationary Series

If $k = 1$, i.e., $X = (x_1,....,x_n)'$, the band-regression estimator (7) can be written as

$$\hat{\beta}_r = \left(\sum_{j=1}^{r} \begin{pmatrix} A_x^2(\omega_j) \\ + B_x^2(\omega_j) \end{pmatrix} \right)^{-1} \sum_{j=1}^{r} \begin{pmatrix} A_x(\omega_j) A_y(\omega_j) \\ + B_x(\omega_j) B_y(\omega_j) \end{pmatrix}$$
$$= \frac{\left[\sum_{j=1}^{r} \begin{pmatrix} (A_x(\omega_j) - iB_x(\omega_j))\overline{(A_y(\omega_j) - iB_y(\omega_j))} \\ + (A_x(\omega_j) + iB_x(\omega_j))\overline{(A_y(\omega_j) + iB_y(\omega_j))} \end{pmatrix} \right]}{2 \sum_{j=1}^{r} \left(A_x^2(\omega_j) + B_x^2(\omega_j) \right)}$$

$$= \frac{\sum_{j=1}^{r}\left(I_{xy}(\omega_j)+I_{xy}(-\omega_j)\right)}{\sum_{j=1}^{r}\left(I_{xx}(\omega_j)+I_{xx}(-\omega_j)\right)} = \frac{\sum_{j=1}^{r}\mathrm{Re}\left(I_{xy}(\omega_j)\right)}{\sum_{j=1}^{r}I_{xx}(\omega_j)}, \quad (17)$$

where

$$A_x(\omega)-iB_x(\omega) = \sum_{t=1}^{n}x_t\cos(\omega t)-i\sum_{t=1}^{n}x_t\sin(\omega t) \quad (18)$$

is the discrete Fourier transform of the series $x_1,....,x_n$ and

$$I_{xy}(\omega) = \frac{1}{2\pi n}(A_x(\omega)-iB_x(\omega))\overline{(A_y(\omega)-iB_y(\omega))} \quad (19)$$

is the cross periodogram of the series $x_1,....,x_n$ and $y_1,....,y_n$.

If the regressor is not deterministic but the two series are rather jointly stationary, the real part of the cross periodogram is an estimator of the cospectrum, the integral of which is just the covariance between x_t and y_t. However, cospectra of economic time series rarely appear to be constant over all frequencies. If the integral of the cospectrum over the interval (ω_{r+1},π) is negligible compared with that over the interval $(0,\omega_{r+1})$ and the goal is to find an appropriate linear predictor of the form $\hat{y}_{n+1} = x_{n+1}\hat{\beta}$, then it seems natural to use $\hat{\beta}_r$ rather than the conventional LS estimator

$$\hat{\beta} = \frac{\sum_{j=1}^{m}\mathrm{Re}\left(I_{xy}(\omega_j)\right)}{\sum_{j=1}^{m}I_{xx}(\omega_j)}, \quad (20)$$

which minimizes the sum of squared errors. Noting that

$$y_t = \frac{2}{n}\sum_{j=1}^{m}\left(A_y(\omega_j)\cos(\omega_j t)+B_y(\omega_j)\sin(\omega_j t)\right) \quad (21)$$

and

$$x_t = \frac{2}{n}\sum_{j=1}^{m}\left(A_x(\omega_j)\cos(\omega_j t)+B_x(\omega_j)\sin(\omega_j t)\right), \quad (22)$$

if both series are mean corrected and n is odd, the sum of squared errors can also be written as

$$\begin{aligned}
&\sum_{t=1}^{n}\left(y_t-\beta x_t\right)^2 \\
&= \frac{2}{n}\sum_{j=1}^{m}\left(A_y^2(\omega_j)+B_y^2(\omega_j)\right)+\frac{2\beta^2}{n}\sum_{j=1}^{m}\left(A_x^2(\omega_j)+B_x^2(\omega_j)\right) \\
&\quad -\frac{4}{n}\beta\sum_{j=1}^{m}\left(A_x(\omega_j)A_y(\omega_j)+B_x(\omega_j)B_y(\omega_j)\right) \\
&= 4\pi\left(\sum_{j=1}^{m}I_{yy}(\omega_j)+\beta^2\sum_{j=1}^{m}I_{xx}(\omega_j)-2\beta\sum_{j=1}^{m}\mathrm{Re}\left(I_{xy}(\omega_j)\right)\right).
\end{aligned} \quad (23)$$

Incorporating the information that the integral of the cospectrum over the interval (ω_{r+1},π) is negligible leads to the estimator

$$\tilde{\beta}_r = \frac{\sum_{j=1}^{r}\mathrm{Re}\left(I_{xy}(\omega_j)\right)}{\sum_{j=1}^{m}I_{xx}(\omega_j)}, \quad (24)$$

which can also be obtained from the band-regression estimator $\hat{\beta}_r$ by shrinkage towards zero, i.e., by multiplication by

$$0 \le \frac{\sum_{j=1}^{r}I_{xx}(\omega_j)}{\sum_{j=1}^{m}I_{xx}(\omega_j)} \le 1. \quad (25)$$

If the focus is on the relationship between the low-frequency components of two series, the use of the band-regression estimator $\hat{\beta}_r$ is appropriate. However, if it is on prediction in the time domain, the whole frequency range will be used. Ideally, the low-frequency components should be multiplied by $\hat{\beta}_r$ and the high-frequency components by zero. Unfortunately, this differentiation is not possible in the time domain. While already the unrestricted estimator $\hat{\beta}$ strikes a balance between the two extremes, the estimator $\tilde{\beta}_r$ is clearly the better choice because it additionally makes use of available extra information.

2.1.3. Degrees-of-Freedom Adjustment

The variance of a stationary process with variance σ_y^2 can be obtained from its spectral density f_y via

$$\sigma_y^2 = \int_{-\pi}^{\pi}f_y(\omega)d\omega, \quad (26)$$

hence we have for large n

$$\sigma_y^2 \sim 2\pi\frac{1}{m}\sum_{j=1}^{m}I_{yy}(\omega_j). \quad (27)$$

The sum of squared residuals

$$\begin{aligned}
S(r) &= \sum_{j=1}^{n}\left(y_t-\tilde{\beta}_r x_t\right)^2 \\
&= 4\pi\left(\sum_{j=1}^{m}I_{yy}(\omega_j)-\frac{\left(\sum_{j=1}^{r}\mathrm{Re}\left(I_{xy}(\omega_j)\right)\right)^2}{\sum_{j=1}^{m}I_{xx}(\omega_j)}-\frac{2\sum_{j=r+1}^{m}\mathrm{Re}\left(I_{xy}(\omega_j)\right)}{\sum_{j=1}^{m}I_{xx}(\omega_j)}\right)
\end{aligned}$$

$$= 4\pi \sum_{j=1}^{m} I_{yy}(\omega_j) \left(1 - \frac{\left(\sum_{j=1}^{r} \mathrm{Re}\left(I_{xy}(\omega_j) \right) \right)^2}{\sum_{j=1}^{m} I_{xx}(\omega_j) \sum_{j=1}^{m} I_{yy}(\omega_j)} - \frac{2 \sum_{j=r+1}^{m} \mathrm{Re}\left(I_{xy}(\omega_j) \right)}{\sum_{j=1}^{m} I_{xx}(\omega_j) \sum_{j=1}^{m} I_{yy}(\omega_j)} \right) \quad (28)$$

is therefore approximately equal to

$$2m\sigma_y^2 \left(1 - \frac{4\pi^2 \left(\sum_{j=1}^{r} \mathrm{Re}\left(I_{xy}(\omega_j) \right) \right)^2}{m^2 \sigma_x^2 \sigma_y^2} \right) \quad (29)$$

when the integral of the cospectrum over the interval (ω_{r+1}, π) is negligible and n is large.

Now assume that the two processes are independent Gaussian white-noise processes. Then

$$\frac{A_x(\omega_1)}{\sigma_x}, \frac{A_y(\omega_1)}{\sigma_y}, \cdots, \frac{A_x(\omega_m)}{\sigma_x}, \frac{A_y(\omega_m)}{\sigma_y} \quad (30)$$

are i.i.d. $N(0, \frac{n}{2})$, which implies that

$$E\left(\sum_{j=1}^{r} \mathrm{Re}\left(I_{xy}(\omega_j) \right) \right)^2$$
$$= \frac{1}{(2\pi n)^2} \sum_{j=1}^{r} \left(EA_x^2(\omega_j) EA_y^2(\omega_j) + EB_x^2(\omega_j) EB_y^2(\omega_j) \right) \quad (31)$$
$$= \frac{\sigma_x^2 \sigma_y^2}{(2\pi n)^2} r \left(\frac{n}{2} \frac{n}{2} + \frac{n}{2} \frac{n}{2} \right)$$
$$= \frac{\sigma_x^2 \sigma_y^2}{4\pi^2} \frac{r}{2}$$

and

$$ES(r) \sim n\sigma_y^2 \left(1 - \frac{\frac{r}{m}}{n} \right). \quad (32)$$

An approximately unbiased estimate of σ_y^2 is given by

$$\sigma_y^2(r) = \frac{\frac{1}{n} \sum_{j=1}^{n} \left(y_t - \tilde{\beta}_r x_t \right)^2}{\left(1 - \frac{\frac{r}{m}}{n} \right)} = \frac{\sum_{j=1}^{n} \left(y_t - \tilde{\beta}_r x_t \right)^2}{n - \frac{r}{m}}, \quad (33)$$

and $n\log(n\sigma_y^2)$ can be estimated by

$$n\log(S(r)) - \log\left(1 - \frac{\frac{r}{m}}{n} \right)^n \sim n\log(S(r)) + \frac{r}{m}, \quad (34)$$

In the standard case, when all frequencies are used, i.e., $r = m$, the number of degrees of freedom is just $n-1$. In contrast, it is a fractional number, namely $n - \frac{r}{m}$, when only $r < m$ frequencies are used.

2.2. Downweighting

Pesaran and Pick [10] investigated the performance of univariate one-step-ahead forecasts of the form

$$\hat{y}_{n+1} = \sum_{j=0}^{n-1} \lambda_{j,n} y_{t-j}, \quad (35)$$

where

$$\sum_{j=0}^{n-1} \lambda_{j,n} = 1, \quad (36)$$

and found that the average-window (*AW*) weights

$$\lambda_{j,n}^{(AW)} = \frac{1}{n}\left(\lambda_{j,j+1}^{(SW)} + \ldots + \lambda_{j,n}^{(SW)} \right) = \frac{1}{n}\left(\frac{1}{j+1} + \ldots + \frac{1}{n} \right) \quad (37)$$

(see [3]), which are obtained by averaging across all subsamples

$$(y_n), (y_n, y_{n-1}), \ldots, (y_n, \ldots, y_{n-j+1}), \\ (y_n, \ldots, y_{n-j}), \ldots, (y_1, \ldots, y_n), \quad (38)$$

outperform the single-window (*SW*) weights

$$\lambda_{j,n}^{(SW)} = \frac{1}{n} \quad (39)$$

unless the breaks are very small. Complementing this result, Steinberger [11] compared these weights in a low-signal/high-noise framework. In his asymptotic analysis, he allowed for an increasing number of breaks. Depending on the specification of the volatility process, the asymptotic mean square forecasting error (MSFE) of the *SW* weights turned out to be asymptotically as small as that of the *AW* weights or even smaller.

Assuming that both the regression parameters and the error variance of a simple regression model are subject to a single structural break, i.e.,

$$Ey_t = \begin{cases} \alpha_1 + \beta_1 x_t & \text{for} \quad t \le t_0 \\ \alpha_2 + \beta_2 x_t & \text{for} \quad t > t_0 \end{cases} \quad (40)$$

and

$$Var(y_t) = \begin{cases} \sigma_1^2 & \text{for} \quad t \le t_0 \\ \sigma_2^2 & \text{for} \quad t > t_0, \end{cases} \quad (41)$$

Pesaran et al. [12] derived optimal weights for one-step-ahead forecasts of the form

$$\hat{y}_{n+1} = \hat{\alpha} + \hat{\beta} x_{n+1}, \quad (42)$$

where

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \left(\sum_{j=0}^{n-1} \lambda_{j,n} \begin{pmatrix} 1 \\ x_{t-j} \end{pmatrix} (1, x_{t-j}) \right)^{-1} \sum_{j=0}^{n-1} \lambda_{j,n} \begin{pmatrix} 1 \\ x_{t-j} \end{pmatrix} y_{t-j}. \quad (43)$$

Their optimal weights have a discrete change which depends on the time and size of the break. This discrete change can only be approximated poorly by weights that

are either constant or slowly decaying. Moreover, the time and size of the structural break are usually unknown and must therefore be estimated. Monte-Carlo results in [12] suggest that the use of estimates instead of the true values for the calculation of the weights is advisable only if the break is large and can easily be identified. Otherwise, a more robust approach is required. Pesaran et al. [12] derived "robust optimal" (*RO*) weights

$$\lambda_{j,n}^{(RO)} \propto \begin{cases} \dfrac{\log(n)}{n-1} \; if \, j = 0 \\[2mm] \dfrac{\log(n)-\log(j)}{n-1} \; if \, j > 0 \end{cases} \qquad (44)$$

by integrating out the effects of uncertainty about the breaks with respect to some given distribution. However, as can be seen from

$$\lambda_{j,k}^{(AW)} = \frac{1}{n}\left(\sum_{k=1}^{n}\frac{1}{k} - \sum_{k=1}^{j}\frac{1}{k}\right) \sim \frac{1}{n}\big(\log(n)-\log(j)\big), \quad (45)$$

the "robust optimal" weights are very similar to the conventional average-window weights, hence it is unrealistic to expect too much of them.

Anyhow, there simply cannot be any fully specified weights that are appropriate for all applications. At least, adjustments to the respective applications must be made with the help of some tuning parameters. For example, an integer-valued tuning parameter can be introduced by requiring a minimum size for the subsamples to be used in window averaging (see [10]) and a real-valued tuning parameter by averaging across exponentially decaying weights rather than equal weights, i.e.,

$$\lambda_{j,n}^{(AE)} = \frac{1}{n}\left(\lambda_{j,j+1}^{(E)} + ... + \lambda_{j,n}^{(E)}\right)$$
$$= \frac{(1-\rho)\rho^{j}}{n}\left(\frac{1}{1-\rho^{j+1}} + ... + \frac{1}{1-\rho^{n}}\right) \qquad (46)$$

where

$$\lambda_{j,n}^{(E)} = \frac{1-\rho}{1-\rho^{n}}\rho^{j} \qquad (47)$$

and

$$\lambda_{j,n}^{(AE)} = \frac{\rho^{j}}{n}\left(\frac{1}{1+\rho+...+\rho^{j}} + ... + \frac{1}{1+\rho+...+\rho^{n-1}}\right) \to \lambda_{j,n}^{(AW)} \quad (48)$$

as $\rho \to 1$. Clearly, it can never be ruled out that in a concrete application some ad-hoc weights such as

$$\lambda_{j,n}^{(AH)} \propto (1+j)^{-\tau} \qquad (49)$$

outperform weights that are supposedly more sophisticated.

3. Empirical Results

Downweighting older observations is an established method to take care of parameter instability in forecasting tasks. Pesaran et al. [12] derived optimal weights for one-step-ahead forecasts under the assumption that information about the location and the size of the breaks is available and then obtained "robust optimal" weights by

integrating out the effects of uncertainty about the breaks with respect to some given distribution. Using the latter weights to forecast real GDP with the slope of the yield curve as explanatory variable, they observed a significant improvement over conventional methods. In the following, we will examine whether the focus on certain frequency bands can lead to a further improvement. For the sake of comparability, the same data set will be used. This data set contains quarterly observations on the real GDP as well as long and short interest rates from 1979Q1 to 2009Q4 of Australia, Canada, France, Germany, Italy, Japan, Spain, UK, and USA. It is part of the larger data set "GVAR data", which is an extension of the data set originally used by Dees et al. [17] and can be downloaded from http://www-cfap.jbs.cam.ac.uk/research/gvartoolbox/download.html.

The slope of the yield curve is supposedly a valuable leading indicator of GDP growth (e.g. [18-23], [12]). Forecasts of the GDP growth rate at time n, y_{n+1}, may therefore be obtained from (42) and (43), where the explanatory variable x_{n+1} is the slope of the yield curve at time n. The growth rates are defined as the differenced log GDP values (multiplied by 100) and the slopes as the differences between the long-term interest rates and the short-term interest rates (yield-curve spreads).

At first glance, forecasts of the form (42) seem a bit naïve because no lagged values of the dependent variable are used. But this is only a serious problem if SW-weights are used in (43). In the case of declining weights, the dummy variable takes over the job of modeling the conditional mean and can compete with low-order ARMA models, at least when there are tuning parameters. For example, the best forecast of a simple invertible MA(1) process

$$y_t = \theta u_{t-1} + u_t \qquad (50)$$

is just given by

$$\hat{y}_{t+1} = \frac{\theta}{1+\theta L}y_t = \theta\sum_{j=0}^{\infty}\theta^{j}y_{t-j}. \qquad (51)$$

It turns out that the dummy variable is quite successful in modeling the conditional mean. Actually, it is so successful that there is hardly anything left to explain for the true explanatory variable, namely the spread.

Following [3] recursive out-of-sample forecasts are constructed for the period from 1994Q1 to 2009Q4. Table 1.a compares averages (over all countries) of sums of squared forecast errors. Consistent with the findings in [3], the forecasts obtained with the *AW* weights and the *RO* weights outperform those obtained with the SW weights. While it is not surprising - because of relationship (45) - that the former two are practically equivalent, it is remarkable that these "optimal" weights are clearly outperformed by some ad-hoc weights such as (49), and to make matters worse, even without fine-tuning, i.e., simply with $\tau = 1$. Even more remarkable is the fact that the performance of the forecasts using the "optimal" weights does not deteriorate when the spread is removed and only the dummy variable is left, which raises a doubt whether the yield-curve spread is really a valuable leading indicator of GDP growth. In contrast, using the lagged GDP growth instead of the spread in addition to the dummy variable immediately boosts the forecasting performance. For a more informative comparison of the

competing forecasts, we use cumulative sums of forecast errors rather than total sums. Figure 1 shows the cumulative sums of absolute forecast errors separately for each country. Despite the fact that absolute errors are used rather than squared errors for reasons of robustness, the curves still fluctuate considerably. Overall, the results are largely consistent across countries and in line with Table 1.a. There are no apparent country-specific characteristics. The only exception is Australia where the benchmark model (with dummy variable plus spread and SW weights) is competitive.
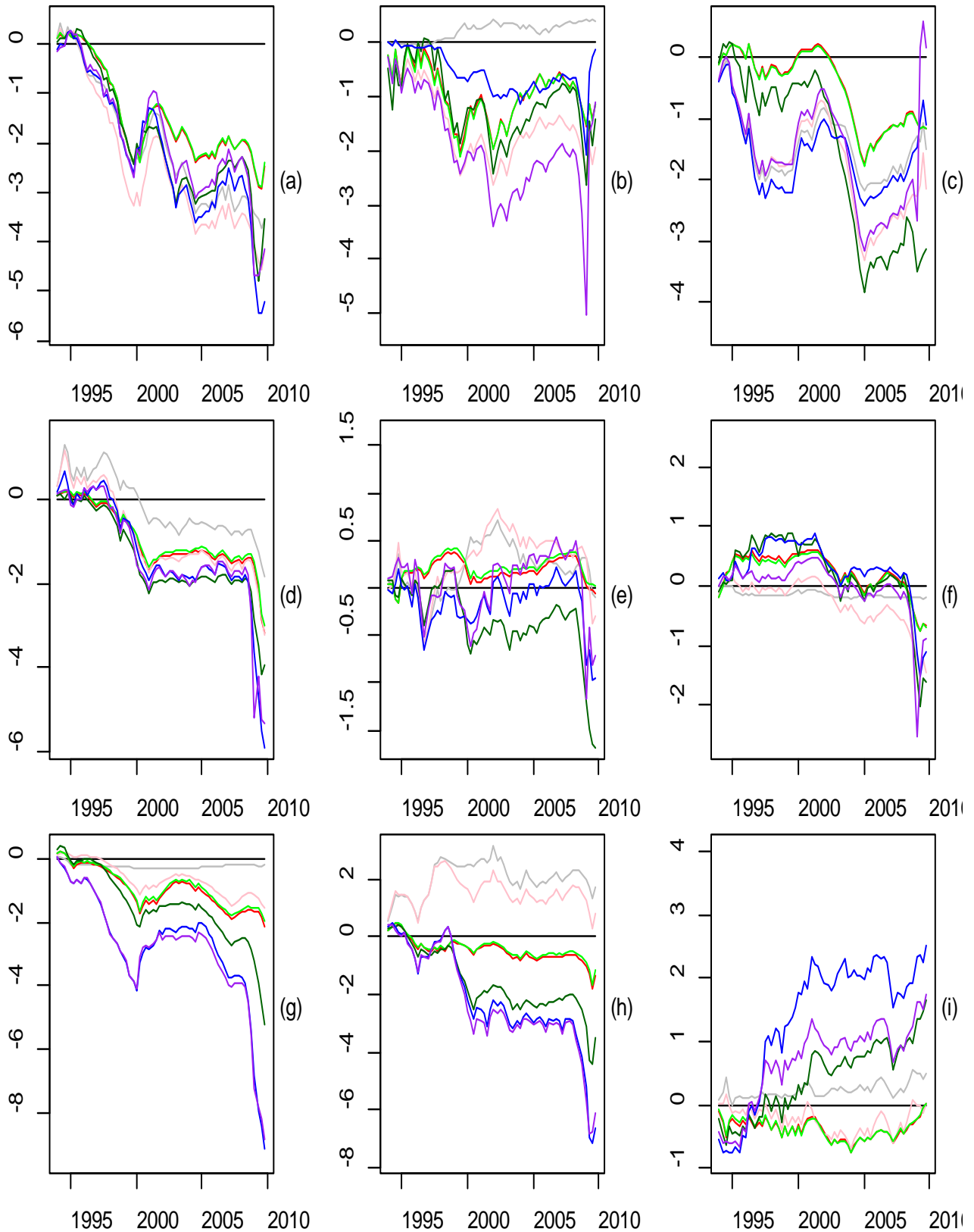

Figure 1. Cumulative absolute forecast errors of models 1 SW (gray), 1+x AW (red), 1 AW (pink), 1+x RO (green), 1+x AH (darkgreen), 1+L(y) SW (blue), 1+L(y) AW (purple) relative to 1+x SW (black): (a) USA (b) Japan (c) Germany (d) UK (e) France (f) Italy (g) Spain (h) Canada (i) Australia

Table 1. Averages (over all countries) of the sums of squared forecast errors of weighted (band) regressionsof GDP growth (*y*) at time *n*on a dummy variable (1) and the yield-curve spread (*x*) at time *n*-1. Also included are forecasts where the spread is replaced by the lagged GDP growth (*Ly*). The conventional forecast $\hat{y}_{n+1} = \hat{\alpha} + \hat{\beta}x_{n+1}$ based on the whole frequency band and obtained with SW weights serves as a benchmark (=1.000)

(a) Weighted regressions

1 (*SW*)	1 (*AW*)	1+*x* (*SW*)	1+*x* (*AW*)	1+*x* (*RO*)	1+*x* (*AH**)	1+*Ly* (*SW*)	1+*Ly* (*AW*)
0.977	0.913	1.000	0.911	0.915	0.795	0.763	0.772

* $\tau = 1$

(b) Weighted band regressions

r/m (the lowest *r* of *m* Fourier frequencies are used)

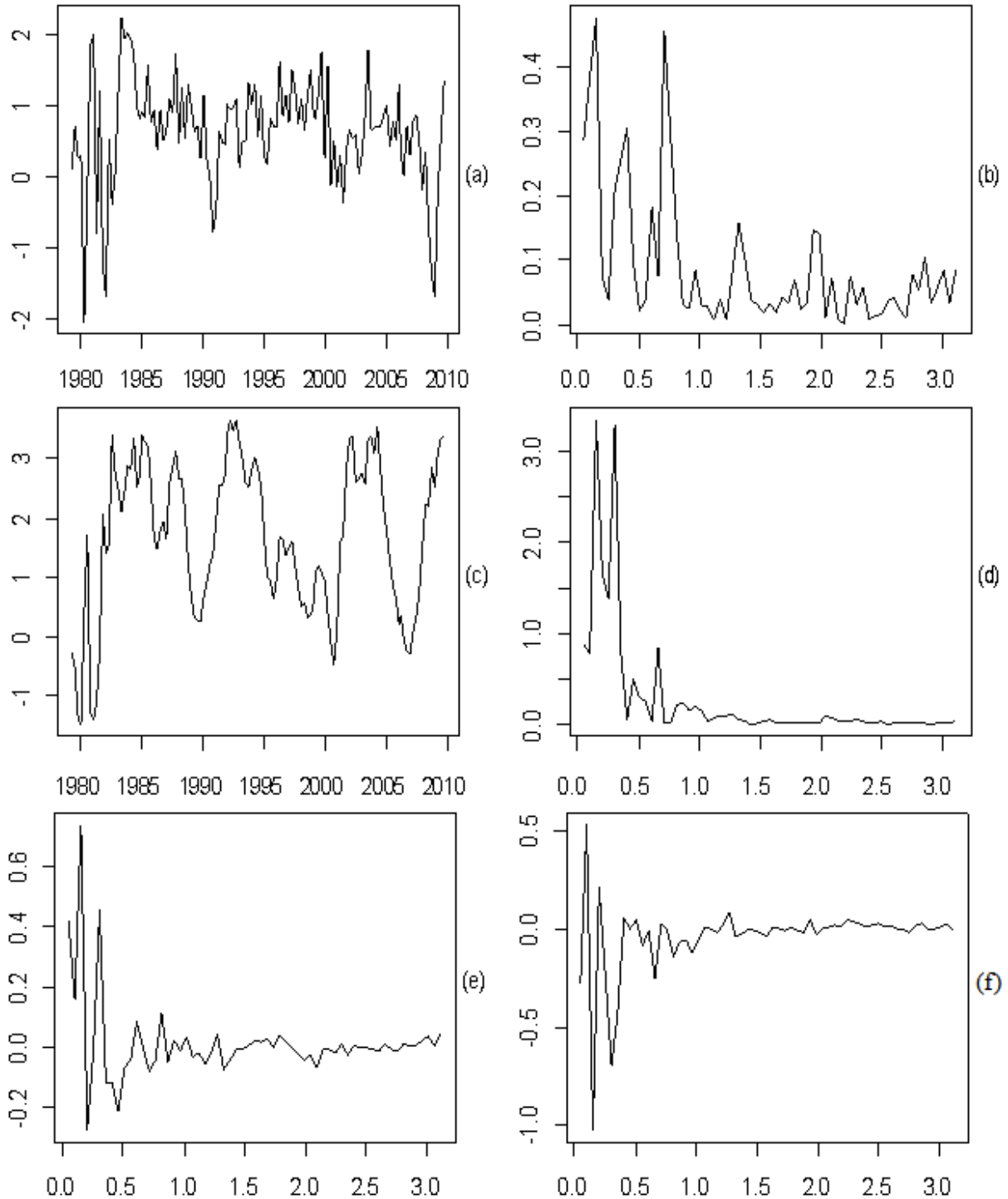|	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50
1+*x* (SW)	0.978	1.000	0.989	0.993	0.999	0.999	0.997	1.001
1+*x* (AW)	0.919	0.911	0.919	0.918	0.917	0.914	0.914	0.913
1+*Ly* (SW	0.865	0.782	0.747	0.737	0.731	0.731	0.732	0.734
1+*Ly* (AW)	0.824	0.771	0.728	0.725	0.732	0.740	0.757	0.771

Figure 2. Frequency-domain analysis of U.S. data: (a) GDP growth, (b) Periodogram of GDP growth, (c) Spread, (d) Periodogram of spread, (e) Co-periodogram (f) Quadrature periodogram

Figure 2 displays the results of a frequency-domain analysis of the U.S. data. The size of the periodogram at a particular frequency indicates how strongly oscillations of this frequency are represented in a time series. The co-periodogram indicates for each frequency how strongly two time series oscillate with a phase difference of zero or half a cycle and the quadrature periodogram indicates how strongly they oscillate with a phase difference of a quarter cycle. Figure 3 displays the co-periodograms of GDP growth and (lagged) spread for all nine countries. Figure 2 and Figure 3 show that the variance of the growth rate, the variance of the spread, and the covariance between the growth rate and the spread are largely determined by the low frequency components of these time series. A band-regression approach may therefore be promising. Thus, the shrinkage estimator (24), which has been derived in the previous section, is also included in the comparative study. Table 1.b shows the results. Again, cumulative sums of absolute forecast errors are displayed separately for each country to provide more detailed information about differences in the forecasting performance (see Figure 4). Apparently, focusing only on some low frequency band has no effect on the forecasts utilizing the spread. They are just as bad as before.
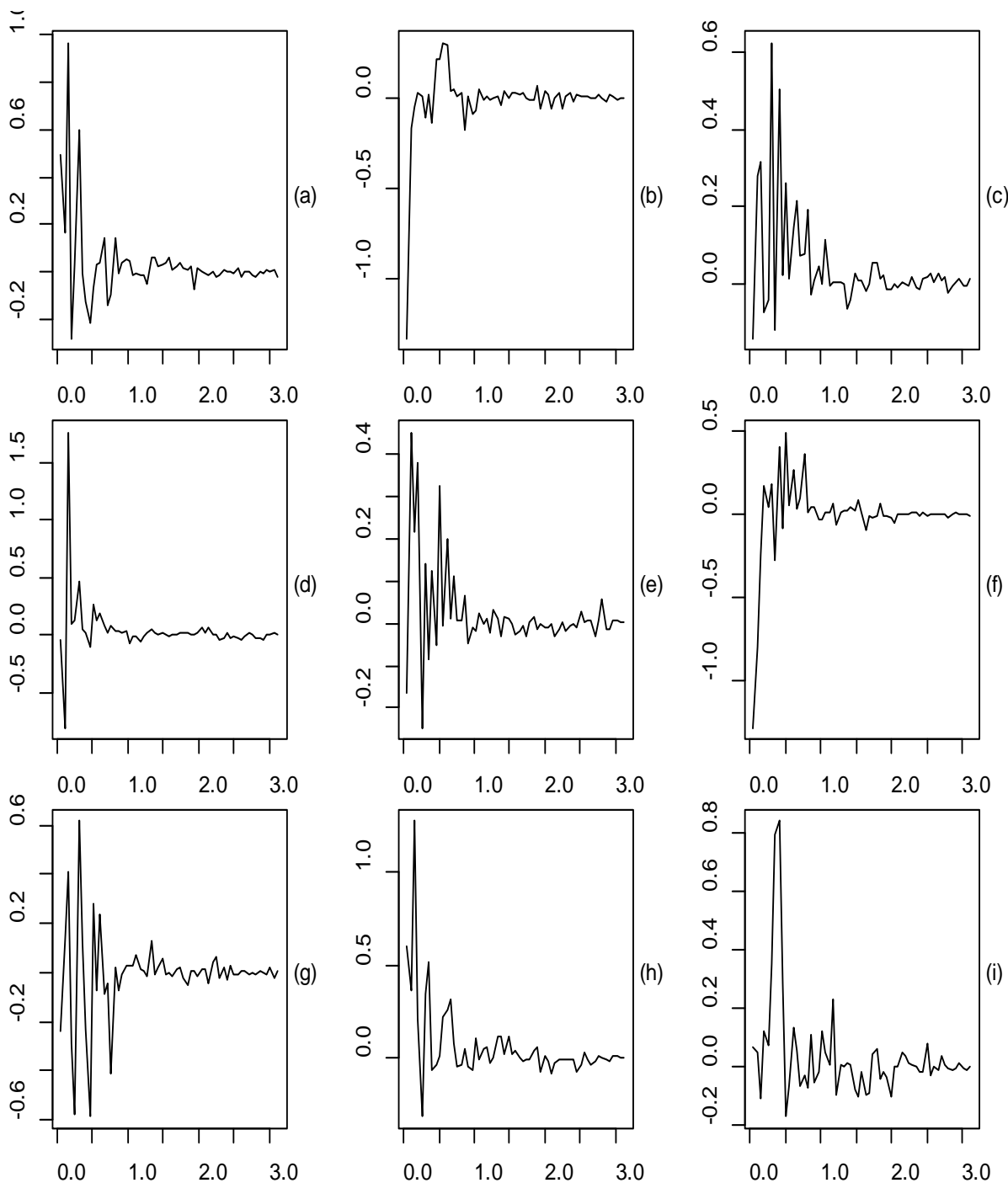

Figure 3. Co-periodogram of GDP growth and lagged spread: (a) USA (b) Japan (c) Germany (d) UK (e) France (f) Italy (g) Spain (h) Canada (i) Australia

However, when the spread is replaced by a variable with predictive power additional to the dummy variable, the band-regression approach works quite well. The best results are obtained when the concepts of downweighting and band regression are combined (see the last line of Table 1.b). A suitable upper limit bound for the frequency

band is in the range from 0.15π and 0.25π. The fractional BIC

$$n \log(S(r)) + (k+1)\frac{r}{m}\log(n) \qquad (52)$$

selects also an upper bound in that range (on an average about 0.15π). In contrast, the fractional AIC

$$n \log(S(r)) + 2(k+1)\frac{r}{m} \qquad (53)$$

is of little help as it typically selects all frequencies. Both the fractional AIC and the fractional BIC are obtained from their conventional counterparts simply by replacing the degrees of freedom k occurring in these criteria by the fractional degrees of freedom kr/m.
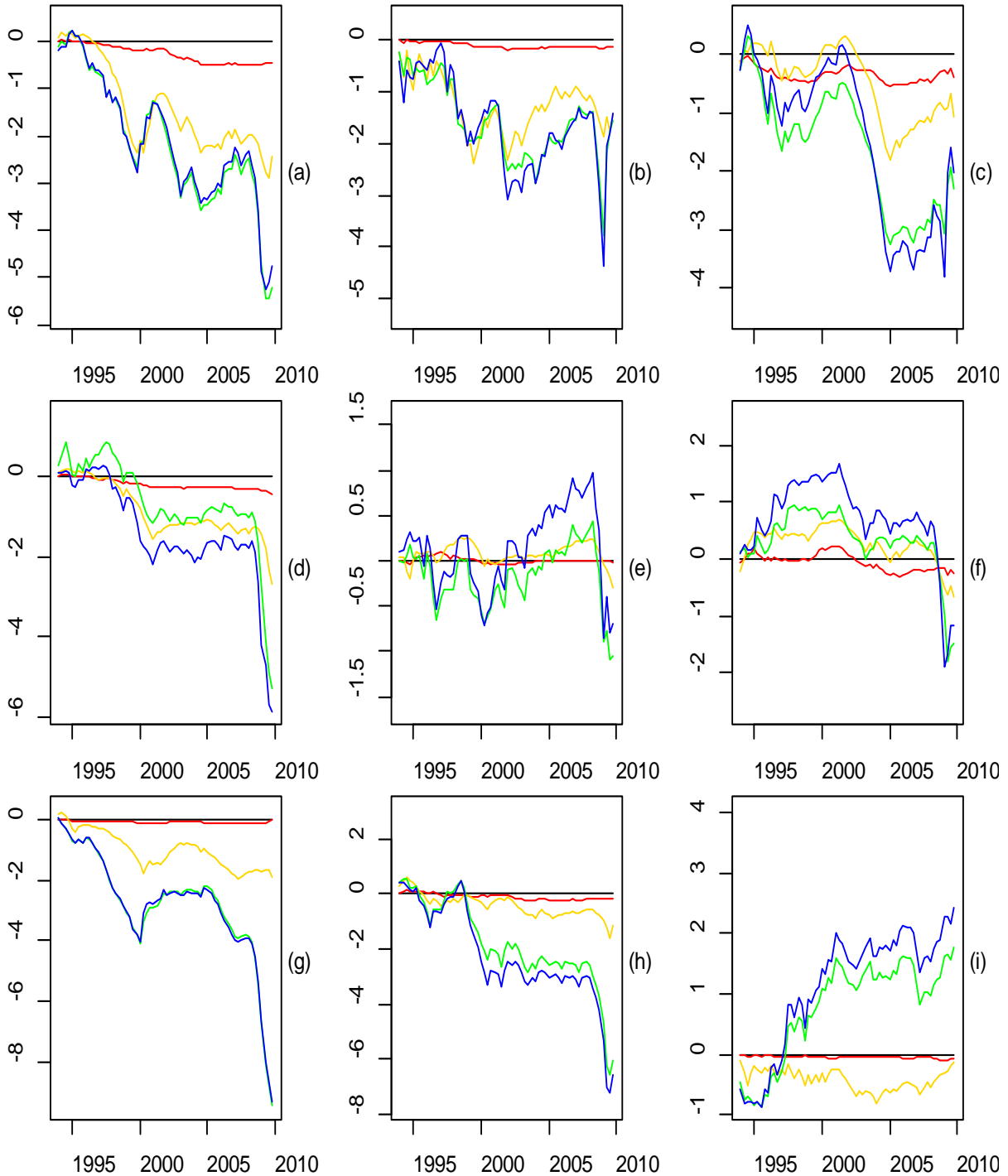

Figure 4. Cumulative absolute forecast errors of band regression ($r=0.2m$) models 1+x SW (red), 1+x AW (gold), 1+L(y) SW (green), 1+L(y) AW (blue) relative to standard regression model 1+x SW (black): (a) USA (b) Japan (c) Germany (d) UK (e) France (f) Italy (g) Spain (h) Canada (i) Australia

4. Discussion

Using frequency-domain information for forecasting purposes is a nontrivial task because it is not clear how the

information contained in the phases should be processed unless there are strict periodicities. In contrast, forecasts based on time-domain models such as ARMA models depend only on the most recent observations and thereby utilize phase information automatically. Of course, estimates of the parameters of time-domain models can

also be obtained indirectly from the amplitudes, but this extra effort will only be worthwhile if there is some advantage over straightforward time-domain estimation. For example, if a linear relationship between variables exists only in a certain frequency band, it will make sense to focus just on this frequency band and ignore the other frequencies. However, the improved estimator of the linear relationship obtained in this way, i.e., the band-regression estimator, is only relevant for certain frequencies and completely irrelevant for the other frequencies. The estimator proposed in this paper therefore forges a compromise between these extremes by shrinking the band-regression estimator towards zero, where the right amount of shrinkage is determined by imposing frequency-domain restrictions on the conventional time-domain estimator.

This estimator is then applied to forecast international GDP growth rates and its forecasting performance is compared with that of other estimators. In addition, the effect of downweighting older observations to take care of structural breaks is examined. The new estimator applied to downweighted observations comes out best. Remarkably, it makes no difference whether or not the slope of the yield-curve is included in the model. Previous evidence that this variable may be a valuable leading indicator of GDP growth is attributed to a possible misspecification of the univariate dynamics of GDP growth.

In principle, the pooling of cross section and time series data would allow the investigation of country specific and time specific factors. However, this is beyond the scope of this paper and is left for future work.

Acknowledgement

References

[1] Barnard, G., "New methods of quality control, "*Journal of the Royal Statistical Society*: *Series A*, 1963, 126, 255-258.

[2] Bates, J.M. and Granger, C.W.J., "The Combination of Forecasts," *OR*, 20, 451-468. 1969.

[3] Pesaran, M.H. and Timmermann, A., "Selection of estimation window in the presence of breaks," *Journal of Econometrics*, 137: 134-161. 2007.

[4] Clark, T.E. and McCracken, M.W., "Improving forecast accuracy by combining recursive and rolling forecasts," *IER*, 50, 363-395. 2009.

[5] Bai, J. and Perron, P., "Estimating and testing linear models with multiple structural changes," *Econometrica*, 66: 47-78. 1998.

[6] Bai, J. and Perron, P., "Computation and analysis of multiple structural change models," *Journal of Applied Econometrics*, 18, 1-22. 2003.

[7] Preinerstorfer, D., "Linear forecasting and subset-selection-based forecasting in structural break models," Master's thesis, University of Vienna, Austria. 2011.

[8] Reschenhofer, E., Preinerstorfer, D. and Steinberger, L., "Non-monotonic penalizing for the number of structural breaks," *Computational Statistics*, 28, 2585-2598. 2013.

[9] Paye, B.S. and Timmermann, A., "Instability of return prediction models," *Journal of Empirical Financ*e, 13, 274-315. 2006.

[10] Pesaran, M.H. and Pick, A., "Forecast combination across estimation windows," *Journal of Business and Economic Statistics*, 29: 307-318. 2011.

[11] Steinberger, L., "Forecasting random walks with structural breaks: averaging across estimation windows," Master's thesis, University of Vienna, Austria. 2012.

[12] Pesaran, M.H., Pick, A. and Pranovich, M., "Optimal forecasts in the presence of structural breaks," *Journal of Econometrics*, 177: 134-152. 2013.

[13] Hannan, E. J., "Regression for Time Series," in *Proceedings of the Symposium on Time Series Analysis*, John Wiley and Sons, 14-37. 1963.

[14] Engle, R. F., "Band spectrum regression, "*International Economic Review*, 15: 1-11. 1974.

[15] Phillips, P.C. B., "Spectral regression for co-integrated time series," in *Nonparametric and Semiparametric Methods in Economics and Statistics*, Cambridge University Press. 1991.

[16] R Core Team, "A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. 2013.

[17] Dees, S., Di Mauro, F., Pesaran, M.H. and Smith, L.V., "Exploring the international linkages of the euro area: a global VAR analysis," *Journal of Applied Econometrics*, 22: 1-38. 2007.

[18] Estrella, A. and Hardouvelis, G.A., "The term structure as a predictor of real economic activity," *Journal of Finance*, 46: 555-576. 1991.

[19] Estrella, A. and Mishkin, F.S., "The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank," *European Economic Review* 41: 1375-1401. 1997.

[20] Stock, J. H. and Watson, M.W., "Forecasting output and inflation: The role of asset prices," *Journal of Economic Literature* 41: 788-829. 2003.

[21] Estrella, A., Rodriguez, A.P. andSchich, S., "How stable is the predictive power of the yield curve? Evidence from Germany and the United States," *Review of Economics and Statistics*, 85: 629-644. 2003.

[22] Giacomini, R. and Rossi, B., "How stable is the forecasting performance of the yield curve for output growth?," *Oxford Bulletin of Economics and Statistics*, 68: 783-795. 2006.

[23] Schrimpf, A. and Wang, Q., "A reappraisal of the leading indicator properties of the yield curve under structural instability," *International Journal of Forecasting*, 26: 836-857. 2010.