Empirical Bayes Forecasts of One Time Series Using Many Predictors

Thomas Knox

Department of Economics, Harvard University

James H. Stock

Kennedy School of Government, Harvard University
and the NBER

and

Mark W. Watson*

Woodrow Wilson School, Princeton University
and the NBER

September 2000

*The authors thank Gary Chamberlain, Ron Gallant, Carl Morris, and Whitney Newey for helpful discussions. This research was supported in part by National Science Foundation grant no. SBR-9730489.

ABSTRACT

We consider both frequentist and empirical Bayes forecasts of a single time series using a linear model with T observations and K orthonormal predictors. The frequentist formulation considers estimators that are equivariant under permutations (reorderings) of the regressors. The empirical Bayes formulation (both parametric and nonparametric) treats the coefficients as i.i.d. and estimates their prior. Asymptotically, when K is proportional to T the empirical Bayes estimator is shown to be: (i) optimal in Robbins' (1955, 1964) sense; (ii) the minimum risk equivariant estimator; and (iii) minimax in both the frequentist and Bayesian problems over a class of nonGaussian error distributions. Also, the asymptotic frequentist risk of the minimum risk equivariant estimator is shown to equal the Bayes risk of the (infeasible subjectivist) Bayes estimator in the Gaussian case, where the "prior" is the weak limit of the empirical cdf of the true parameter values. Monte Carlo results are encouraging. The new estimators are used to forecast monthly postwar U.S. macroeconomic time series using the first 151 principal components from a large panel of predictors.

Key Words: Large model forecasting, equivariant estimation, minimax forecasting

JEL Codes: C32, E37

1. Introduction

Recent advances in data availability now allow economic forecasters to examine hundreds of economic time series during each forecasting cycle. Consider, for example, the problem of forecasting of the U.S. index of industrial production. A forecaster can collect monthly observations on, say, 200 (or more) potentially useful predictors beginning in January 1959. But what should the forecaster do next?

This paper considers this problem for a forecaster using a linear regression model with K regressors and T observations, whose loss function is quadratic in the forecast error. The regressors are taken to be orthonormal, an assumption that both simplifies the analysis and is motivated by the empirical work summarized in section 6, in which the regressors are principal components of a large macroeconomic data set. In this framework, the forecast risk (the expected value of the forecast loss) is the sum of two parts: one that reflects unknowable future events, and one that depends on the estimator used to construct the forecast. Because the forecaster can do nothing about the first, we focus on the second part, which is the estimation risk. Because the regressors are orthonormal, we take this to be $\mathrm{tr}(V_{\tilde{b}})$, the trace of the mean squared error matrix of the candidate estimator, $\tilde{b}$.

This econometric forecasting problem thus reduces to the statistical problem of finding the estimator that minimizes $\mathrm{tr}(V_{\tilde{b}})$. When K is large, this (and the related K-mean problem) is a difficult problem that has received much attention ever since Stein (1955) showed that the ordinary least squares (OLS) estimator is inadmissible for $K \geq 3$. A variety of approaches are available in the literature, including model selection, model averaging, shrinkage estimation, ridge regression, and parameter reduction schemes such as factor models (references are given below). However, outside of a subjectivist Bayesian framework (where the optimal estimator is the posterior mean), the quest for an optimal estimator has been elusive.[1]

We attack this problem from two perspectives, one classical ("frequentist"), and one Bayesian.

Our frequentist approach to this problem is based on the theory of equivariant estimation. Suppose for the moment that the regression errors are i.i.d. normally distributed, that they are independent of the regressors, and that the regressor and error distributions do not depend on the regression parameters; this shall henceforth be referred to as the "Gaussian case". In the Gaussian case, the likelihood does not depend on the ordering of the regressors, that is, the likelihood is invariant to simultaneous permutations of the cross sectional index of the regressors and their coefficients. Moreover, in the Gaussian case the OLS estimators are sufficient for the regression parameters. These two observations lead us to consider the class of estimators that are equivariant functions of the OLS estimator under permutations of the cross sectional index. Because the form of these estimators derives from the Gaussian case, we call these "Gaussian equivariant estimators." This class is large, and contains common estimators in this problem, including OLS, OLS with information criterion selection, ridge regression, the James-Stein (1960) estimator, the positive part James-Stein estimator, and common shrinkage estimators. The estimator that minimizes tr($V_{\tilde{b}}$) among all equivariant estimators is the minimum risk equivariant estimator. If this estimator

[1]Two other approaches to the many-regressor problem that have attracted considerable attention are Bayesian model selection and Bayesian model averaging. Recent developments in Bayesian model averaging are reviewed by Hoeting, Madigan, Raftery and Volinsky (1999). Some recent developments in Bayesian model selection are reviewed by George (1999). The work in this literature that is, as far as we know, closest to the present paper is George and Foster (2000), which considers an empirical Bayes approach to variable selection. However, their setup is fully parametric and their results refer to model selection, a different objective than ours.

achieves the minimum risk uniformly for all true regression coefficients in an arbitrary closed ball around the origin, the estimator is uniformly minimum risk equivariant over this ball.

The Bayesian approach has two different motivations. One is to adopt the perspective of a subjectivist Bayesian, and to model the coefficients as i.i.d. draws from some subjective prior distribution G_b. This leads to considering the Bayes risk, $\int \mathrm{tr}(V_{\tilde{b}}) \mathrm{d}G_b(b_1) \cdots \mathrm{d}G_b(b_K)$, rather than the frequentist risk $\mathrm{tr}(V_{\tilde{b}})$. The subjectivist Bayesian knows his prior, and because of quadratic loss the Bayes estimator is the posterior mean. In the Gaussian case, this depends only on the OLS coefficients, and computation of the subjectivist Bayes estimator is straightforward. A different motivation is to adopt an empirical Bayes perspective and to treat the "prior" G as an unknown infinite dimensional nuisance parameter. Accordingly, the empirical Bayes estimator is the subjectivist Bayes estimator, constructed using an estimate of G.[2] We adopt this latter perspective, and consider empirical Bayes estimators of b. In the Gaussian case, the OLS estimators are sufficient for the regression parameters, so we consider empirical Bayes estimators that are functions of the OLS estimators. In parallel to the nomenclature in the frequentist approach, we refer to these as "Gaussian empirical Bayes" estimators.

Although the form of these estimators is motivated by the Gaussian case, the statistical properties of these estimators are examined under more general conditions on the joint distribution of the regression errors and regressors, such as existence of certain moments, smoothness of certain distributions, and mixing. Accordingly, all our results are asymptotic. If K is held fixed as $T \to \infty$, the risk of all mean-square consistent estimators converges to zero, and the forecaster achieves the

---

[2] Empirical Bayes methods were introduced by Robbins (1955, 1964). Efron and Morris (1972) showed that the James-Stein estimator can be derived as an empirical Bayes estimator. Maritz and Lwin (1989) and Carlin and Louis (1996) provide a recent monograph treatments of empirical Bayes methods.

optimal forecast risk for any such estimator. This setup does not do justice to the forecasting problem with, say, K = 200 and T = 500. We therefore adopt a nesting that treats K as proportional to T (an assumption used, in a different context, by Bekker [1994]); specifically, $K/T \to r$ as $T \to \infty$. If the true regression coefficients are fixed, then as K increases the population R^2 of the forecasting regression approaches one. This also is unrealistic, so for the asymptotic analysis we model the true coefficients as being in a $1/\sqrt{T}$ neighborhood of zero. Under this nesting, the estimation risk (frequentist and Bayesian) has a nontrivial (nonzero but finite) asymptotic limit.

This paper makes three main theoretical contributions.

The first concerns the Bayes risk. In the Gaussian case, we show that a Gaussian empirical Bayes estimator asymptotically achieves the same Bayes risk as the subjectivist Bayes estimator, which treats G as known. This is shown both in a nonparametric framework, in which G is treated as an infinite dimensional nuisance parameter, and in a parametric framework, in which G is finitely parameterized. Thus this Gaussian empirical Bayes estimator is asymptotically optimal in the Gaussian case in the sense of Robbins (1964), and the Gaussian empirical Bayes estimator is admissible asymptotically. Moreover, the same Bayes risk is attained under the weaker, non-Gaussian assumptions on the distribution of the error term and regressors. Thus, the Gaussian empirical Bayes estimator is minimax (as measured by the Bayes risk) against a large class of distributional deviations from the assumptions of the Gaussian case.

The second contribution concerns the frequentist risk. In the Gaussian case, the Gaussian empirical Bayes estimator is shown to be, asymptotically, the uniformly minimum risk equivariant estimator. Moreover, the same frequentist risk is attained under weaker, non-Gaussian assumptions. Thus, the Gaussian empirical Bayes estimator is minimax (as measured by the frequentist risk) among equivariant estimators against these deviations from the Gaussian case.

Third, because the same estimator solves both the Bayes and the frequentist problems, it makes sense that the problems themselves are the same asymptotically. We show that this is so. Specifically, it is shown that the empirical Bayes estimator asymptotically achieves the same Bayes risk as the subjectivist Bayes estimator based on the "prior" which is the weak limit of the cdf of the true regression coefficients (assuming this exists). Furthermore, this Bayes risk equals the limiting frequentist risk of the minimum risk equivariant estimator.

This paper also makes several contributions within the context of the empirical Bayes literature. Although we do not have repeated forecasting experiments, under our asymptotic nesting in the Gaussian case the regression problem becomes formally similar to the Gaussian compound decision problem. Also, results for the compound decision problem are extended to the nonGaussian case by exploiting Berry-Esseen type results for the regression coefficients; this leads to our minimax results. Finally, permutation arguments are used to extend an insight of Edelman (1988) in the Gaussian compound decision problem to show that the empirical Bayes estimator is also minimum risk equivariant.

The remainder of the paper is organized as follows. The model, Bayesian risk function, and Gaussian empirical Bayes estimators are presented in section 2. Assumptions and theoretical results regarding the OLS estimators and the Bayes risk are given in section 3. The frequentist equivariant estimation problem is laid out in section 4, and the minimum risk equivariant estimator is characterized in section 5. The link between the two problems is discussed in section 6. Section 7 summarizes a Monte Carlo study of the Gaussian empirical Bayes estimators from both Bayesian and frequentist perspectives. An empirical application, in which these methods are used to forecast several U.S. macroeconomic time series, is summarized in section 8. Section 9 concludes.

2. The Model, Bayes Risk, and Gaussian Empirical Bayes Estimators

2.1. The Model and Asymptotic Nesting

We consider the linear regression model,

$$(2.1) \qquad y_{t+1} = b'X_t + e_{t+1},$$

where X_t is a vector of K predictor time series and e_{t+1} is an error term that is a homoskedastic martingale difference sequence with $E[e_{t+1}|F_t] = 0$ and $E(e_{t+1}^2|F_t) = s_e^2$, where $F_t = \{X_t, e_t, X_{t-1}, e_{t-1},\ldots\}$.

We are interested in out-of-sample forecasting, specifically forecasting y_{T+1} using X_T under quadratic loss. Let $\tilde{b}$ be an estimator of b constructed using observations on $\{X_{t-1}, y_t, t=1,\ldots,T\}$, and let $\tilde{y}_{T+1} = \tilde{b}'X_T$ be a candidate forecast of y_{T+1}. The forecast loss is,

$$(2.2) \qquad L(\tilde{b}, b) = (y_{T+1} - \tilde{y}_{T+1})^2 = [e_{T+1} + (\tilde{b} - b)'X_T]^2.$$

The forecast risk is the expected loss,

$$(2.3) \qquad EL(\tilde{b}, b) = s_e^2 + E(\tilde{b} - b)'H_T(\tilde{b} - b),$$

where $H_T = E(X_T X_T' \mid F_{T-1}, y_T)$.

To keep the analysis tractable, we make two simplifying assumptions: first, that the regressors are orthonormal, so $T^{-1}\sum_{t=1}^{T} X_t X_t' = I_K$; and second, that $H_T = I_K$, the $K\times K$ identity matrix. As discussed in the introduction, the first assumption arises from our empirical application, in which the regressors are orthonormal by construction. The second assumption, that $H_T = I_K$, is motivated by the first: because the regressors are orthonormal, it is plausible to weight each diagonal element of the estimation risk equally and to apply zero weight to the off diagonal elements, that is, to set $H_T = I_K$.

The forecaster would like to minimize the forecasting risk (2.3). Because s_e^2 does not depend on the estimator, only the second term in the forecasting risk is affected by statistical considerations; this is the estimation risk which, upon setting setting $H_T = I_K$, is tr($V_{\tilde{b}}$), where $V_{\tilde{b}} = E(\tilde{b} - b)(\tilde{b} - b)'$. Looking ahead to the asymptotic analysis, we rewrite this term using the change of variables,

$$b = b/\sqrt{T} \text{ and } \tilde{b} = \tilde{b}/\sqrt{T}\ . \tag{2.4}$$

Using this change of variables and setting $H_T = I_K$, the estimation risk tr($V_{\tilde{b}}$) is,

$$R(b,\ \tilde{b}\,) = \left(\tfrac{K}{T}\right)K^{-1}\sum_{i=1}^{K} E(\tilde{b}_i - b_i)^2\ . \tag{2.5}$$

where b_i is the i^{th} element of b, etc.

2.2 Asymptotic Nesting

The asymptotic nesting formalizes the notion that number of regressors is large, specifically, that

(2.6) $$K/T \to r \text{ as } T \to \infty.$$

To simplify notation, we ignore integer constraints and set $K = rT$.

Under this nesting, if each b_i is bounded away from zero, the population R^2 tends to one, which is unrepresentative of the empirical problems of interest. We therefore adopt a nesting in which each predictor is treated as making a small but potentially nonzero contribution to the forecast, specifically we adopt the local parameterization (2.4) with $\{b_i\}$ held fixed as $T \to \infty$. Because K and T are linked, various objects are doubly indexed arrays, and b and its estimates are sequences indexed by K, but to simplify notation this indexing is usually suppressed. All limits in this paper are taken under (2.4) and (2.6).

2.3 OLS Estimators

Under the nesting (2.4) and (2.6) with $T^{-1}\sum_{t=1}^{T} X_t X_t^{'} = I_K$, the OLS estimator of b is,

(2.7) $$\hat{b} = T^{-1/2}\sum_{t=1}^{T} X_{t-1} y_t ,$$

so that $E\hat{b}_i = b_i$ and $E(\hat{b} - b)(\hat{b} - b)' = s_e^2 I_K$. Also let $\hat{s}_e^2$ be the usual estimator of s_e^2,

(2.8) $$\hat{s}_e^2 = (T-k)^{-1}\sum_{t=1}^{T}(y_t - \hat{b}' X_{t-1})^2 .$$

2.4 Class of Estimators and the Bayes Risk

As discussed in the introduction, the estimators we consider are motivated by supposing that e_t is iid N(0, s_e^2), $\{X_t\}$ and $\{e_t\}$ are independent, the distribution of $\{X_t, e_t\}$ does not depend on b, and the distribution of $\{X_t\}$ does not depend on s_e^2; these assumptions constitute the "Gaussian case". In the Gaussian case, ($\hat{b}$, $\hat{s}_e^2$) are sufficient for (b, s_e^2). We therefore restrict attention to estimators $\tilde{b}$ that are functions of ($\hat{b}$, $\hat{s}_e^2$).

The likelihood of $\hat{b}$ (given b and s_e^2), denoted by $f_K(\hat{b}\,|b)$, has the location form, $f_K(\hat{b}\,|b) = f_K(\hat{b} - b)$. In the Gaussian case, $\hat{b}$ has a finite sample N(b, $s_e^2 I_K$) distribution, and it is useful to adopt special notation for this case. Let $f_K(z) = \prod_{i=1}^{K} f(z_i)$, where f is the normal density with mean zero and variance s_e^2. Thus, in the Gaussian case $f_K = f_K$. Note that, if K were fixed, standard central limit theory would imply that f_K provides a good approximation to f_K.

In the Bayesian formulation, $\{b_i\}$ are modeled as i.i.d. draws from the prior distribution G. We suppose that the subjectivist Bayesian knows s_e^2 (has a point mass prior on s_e^2). (One motivation for this simplification is that it will be shown that $\hat{s}_e^2$ is L_2-consistent for s_e^2, so that a proper prior over s_e^2 would be dominated by the information in $\hat{s}_e^2$.) Accordingly, the Bayes risk we consider is the frequentist risk (2.5), integrated against the prior distribution G. Upon setting $K/T = r$, under the normalization (2.4) the Bayes risk is,

(2.9)
$$\begin{aligned} r_G(\tilde{b}, f_K) &= r\int K^{-1}\sum_{i=1}^{K} E(\tilde{b}_i - b_i)^2 \mathrm{d}G_K(b) \\ &= r\int K^{-1}\sum_{i=1}^{K}\int(\tilde{b}_i(\hat{b}) - b_i)^2 f_K(\hat{b} - b)\mathrm{d}\hat{b}\mathrm{d}G_K(b) \end{aligned}$$

where $G_K(b) = G(b_1)\cdots G(b_K)$ and where the second line makes the integration in the inner expectation explicit.

2.5. The Subjectivist Bayes Estimator in the Gaussian Case

The subjectivist Bayesian knows G. Because loss is quadratic, the Bayes risk (2.9) is minimized by the posterior mean. In the Gaussian case, $\{\hat{b}_i\}$ are i.i.d., so the posterior mean is,

(2.10) $$\frac{\int x f_K(\hat{b}-x)\mathrm{d}G_K(x)}{\int f_K(\hat{b}-x)\mathrm{d}G_K(x)}.$$

Because the likelihood is Gaussian, the posterior mean (2.10) can be written in terms of the score of the marginal distribution of $\hat{b}$ (e.g. see Maritz and Lwin [1989, p. 73]). Let m_f denote the marginal distribution of $(\hat{b}_1,...,\hat{b}_K)$ in the Gaussian case,

(2.11) $$m_f(x) = \int f_K(x-b)dG_K(b),$$

and let $\ell_f(x) = m_f'(x)/m_f(x)$ be its score. Accordingly, the subjectivist Bayes estimator $\hat{b}^{NB}$ (the "Normal Bayes" estimator) is,

(2.12) $$\hat{b}^{NB} = \hat{b} + s_e^2\ell_f(\hat{b}).$$

Note that $\hat{b}^{NB}$ is based on the prior G through the score, although this dependence is suppressed notationally.

2.6 Gaussian Empirical Bayes Estimators

The Gaussian empirical Bayes estimators studied here are motivated by the foregoing expressions derived for the Gaussian case. In the empirical Bayes approach to this problem, G is unknown, as is s_e^2. Thus the score ℓ_f is unknown, and the estimator (2.12) is infeasible. However, both the score and s_e^2 can be estimated. Moreover, although the derivation of (2.12) relies on f_K = f_K, as was mentioned f_K might be a plausible large sample approximation to f_K. This suggests using an estimator of the form (2.12), with ℓ_f and s_e^2 estimated, even outside the Gaussian case.

The resulting estimator is referred to as a simple empirical Bayes estimator ("simple" because G does not appear explicitly in (2.12), as it does in (2.10)). Throughout, s_e^2 is estimated by $\hat{s}_e^2$, defined in (2.8). Both parametric and nonparametric approaches to estimating the score are considered. These respectively yield parametric and nonparametric empirical Bayes estimators.

Parametric Gaussian empirical Bayes estimator. The parametric Gaussian empirical Bayes estimator is based on adopting a parametric specification for G, which will be denoted $G_K(b;\theta)$, where θ is a finite dimensional parameter vector. Using the normal approximation f_K to the likelihood, this in turn provides a parametric approximation to the marginal distribution of $\hat{b}$, $m_f(x;\theta) = \int f_K(x-b)dG_K(b;q)$. The parametric Gaussian empirical Bayes estimator is computed by substituting estimates of s_e^2 and q into $m_f(x;\theta)$, using this parametrically estimated marginal and its derivative to estimate the score, and substituting this parametric score estimator into (2.12). The specific parametric score estimator used here is,

$$\hat{\ell}_K(x;\hat{q}) = \frac{m'_{\hat{f}}(x;\hat{q})}{m_{\hat{f}}(x;\hat{q}) + s_K}, \tag{2.13}$$

where $m_{\hat{f}}(x;\hat{q}) = \int \hat{f}_K(x-b)dG_K(b;\hat{q})$ and $m'_{\hat{f}}(x;\hat{q}) = \int \hat{f}'_K(x-b)dG_K(b;\hat{q})$, where $\hat{f}_K$ denotes f_K with s_e^2 estimated, that is, $\hat{f}_K(u) = (2p\hat{s}_e^2)^{-K/2}\exp(-u'u/2\hat{s}_e^2)$, and where $\{s_K\}$ is a sequence of small positive numbers such that $s_K \to 0$. (The sequence $\{s_K\}$ is a technical device used in the proof to control the rate of convergence.)

The parametric Gaussian empirical Bayes estimator, $\hat{b}^{PEB}$, is obtained by combining (2.12) and (2.13) and using $\hat{s}_e^2$; thus,

$$\hat{b}^{PEB} = \hat{b} + \hat{s}_e^2\hat{\ell}_K(\hat{b};\hat{q}). \tag{2.14}$$

Nonparametric Gaussian simple empirical Bayes estimator. The nonparametric Gaussian simple empirical Bayes estimator is based on the assumption that $\{b_i\}$ are independent draws from the distribution G. This permits estimation of the score without assuming a parametric form for G.

The score estimator used for the theoretical results uses a construction similar to Bickel et. al. (1993) and van der Vaart (1988). Let $w(z)$ be a symmetric bounded kernel with $\int z^4 w(z)dz < \infty$ and with bounded derivative $w'(z) = \mathrm{d}w(z)/\mathrm{d}z$, and let h_K denote the kernel bandwidth sequence. Define

(2.15) $$\hat{m}_{iK}(x) = \frac{1}{(K-1)h_K} \sum_{j \neq i} w\left(\frac{\hat{b}_j - x}{h_K}\right)$$

(2.16) $$\hat{m}'_{iK}(x) = -\frac{1}{(K-1)h_K^2} \sum_{j \neq i} w'\left(\frac{\hat{b}_j - x}{h_K}\right), \text{ and}$$

(2.17) $$\breve{\ell}_{iK}(x) = \frac{\hat{m}'_{iK}(x)}{\hat{m}_{iK}(x) + s_K}\,.$$

The nonparametric score estimator considered here is,

(2.18) $$\hat{\ell}_{iK}(x) = \left\{ \begin{matrix} \breve{\ell}_{iK}(x) \text{ if } |x| < \sqrt{\frac{\hat{s}_e^2}{128}\log K} \text{ and } |\breve{\ell}_{iK}(x)| \leq q_K, \\ 0 \text{ otherwise} \end{matrix} \right\},$$

where $\{s_K\}$ and $\{q_K\}$ are sequences of constants. Rates for these sequences are given below.

The nonparametric Gaussian simple empirical Bayes estimator, $\hat{b}^{NSEB}$, obtains by substituting $\hat{s}_e^2$ and (2.18) into (2.12); thus,

(2.19) $$\hat{b}_i^{NSEB} = \hat{b}_i + \hat{s}_e^2 \hat{\ell}_{iK}(\hat{b}_i),\ \ i=1,\ldots,K.$$

Note that although the assumption of normal errors was used to motivate the form of this estimator, the normal approximation f_K is in fact not used in the construction of (2.18).

3. Results Concerning OLS and the Bayes Risk

3.1. Assumptions

The following assumptions are used for one or more of the theoretical results. Throughout, we adopt the notation that C is a finite constant, possibly different in each occurrence.

The first assumption restricts the moments of $\{X_t\}$ and $\{e_t\}$. Let $Z_t = (X_{1t},..., X_{Kt}, e_t)$.

Assumption 1 (moments).

(i) $T^{-1}\sum_{t=1}^{T} X_t X_t' = I_K$;

(ii) $\sup_{it} EX_{it}^{12} \leq C < \infty$ and $\sup_t Ee_t^{12} \leq C < \infty$;

(iii) $E(e_t|Z_{t-1},..., Z_1) = 0$;

(iv) $E(e_t^2|X_t, Z_{t-1},..., Z_1) = s_e^2 > 0$; and

(v) $\sup_T \max_{t\leq T} \sup_{Z_1,...,Z_{t-1}} E(e_t^4 \mid X_t, Z_{t-1},..., Z_1) \leq C < \infty$.

The next assumption is that the maximal correlation coefficient of Z decays geometrically (cf. Hall and Heyde [1980], p. 276). Let $\{n_n\}$ denote the maximal correlation coefficients of Z, that is, let $\{n_n\}$ be a sequence such that $\sup_m \sup_{y\in\mathcal{H}_1^m, x\in\mathcal{H}_{m+n}^\infty} |\mathrm{corr}(x, y)| \leq n_n$, where $\mathcal{H}_a^b$ is the s-field generated by the random variables $\{Z_s, s=a,\ldots,b\}$.

Assumption 2 (time series dependence). $\{Z_t, t=1,\ldots,T\}$ has maximal correlation coefficient n_n that satisfies $n_n \leq De^{-ln}$ for some positive finite constants D and l.

The next assumption places smoothness restrictions on the (conditional) densities of $\{X_{it}\}$ and $\{e_{it}\}$. Let $p^X_{itK}(x)$ denote the conditional density of X_{it} given $(Z_1,\ldots,Z_{t-1})$; let $p^X_{ijtK}(x_i \mid x_j)$ be the conditional density of X_{it} given $(X_{jt}, Z_1,\ldots,Z_{t-1})$; and let $p^e_{itK}(e)$ denote the conditional density of e_t given $(Z_1,\ldots,Z_{t-1})$.

Assumption 3 (densities).

(i) The distribution of $\{X_{it}, e_t\}$ does not depend on $\{b_i\}$.

(ii) There exists a constant $C < \infty$ such that, for all i, j, t, K, $| p^X_{itK}(x) | \le C$, $| p^X_{ijtK}(x_i \mid x_j) | \le C$ for $i \ne j$, and $\int \left| \frac{d^2}{de^2} p^e_{itK}(e) \right| de \le C$.

The next assumption restricts the cross sectional dependence among $\{X_{it}\}$ using a conditional maximal correlation coefficient condition. Let $\underline{X}_i = (X_{i1},\ldots, X_{iT})$ and let $\mathcal{F}^b_a$ be the s-field generated by the random variables $\{\underline{X}_i,\ i = a,\ldots,b\}$, and define the cross sectional conditional maximal correlation coefficients $\{t_n\}$ to be a sequence satisfying $\sup_m \sup_{y \in \mathcal{F}_1^m, x \in \mathcal{F}^\infty_{m+n}} |\mathrm{corr}(x,y|\underline{X}_j)| \le t_n$ for all j.

Assumption 4 (cross sectional dependence). There exists a sequence of cross sectional maximal correlation coefficients $\{t_n\}$ such that $\sum_{n=1}^{\infty} t_n < \infty$.

The next two assumptions place restrictions on the family containing the true distribution of b in, respectively, the parametric and nonparametric cases.

Assumption 5 (parametric G).

(i) $\{b_i\}$ are i.i.d. with distribution G and var(b_i) = $s_b^2 < \infty$.

(ii) G belongs to a known family of distributions $G(b;q)$ indexed by the finite-dimensional parameter q contained in a compact Euclidean parameter space Θ;

(iii) G has density $g(b;\, q)$ which is Lipschitz continuous in q uniformly over b and q, i.e. $\sup_{b,q\in\Theta}\sup_{q'\in\Theta}|g(b;q) - g(b;q')| \le C\|q - q'\|$.

(iv) There exists an estimator $\hat{q} = q(\hat{b}, \hat{s}_e^2)$ such that, for all K sufficiently large, $E[K\|\hat{q} - q\|^2] \le C < \infty$, where the expectation is taken either under (f_K, G) or under (f_K, G).

Assumption 6 (nonparametric G). $\{b_i\}$ are i.i.d. with distribution G and var(b_i) = $s_b^2 < \infty$.

The final assumption provides conditions on the rates of the various sequences of constants used to construct the estimators.

Assumption 7 (rates). The sequences $\{s_K\}$, $\{q_K\}$, and $\{h_K\}$ satisfy: $h_K \to 0$, $s_K \to 0$, $q_K \to \infty$, $K^{1/24}h_K\log K \to 0$, $K^{2/9}h_K \to \infty$, $s_K^2\log K \to \infty$, $K^{-1/6}q_K \to \infty$, and $K^{-1/2}q_K \to 0$.

3.2. Results for the OLS Estimators

The theoretical results pertain to the model (2.1), and all limits are taken under the asymptotic nesting (2.4) and (2.6). The first result is that, under this nesting, $\hat{s}_e^2$ is consistent.

Theorem 1 (standard error of the regression).

Under assumptions 1 and 2, E[($\hat{s}_e^2$ - s_e^2)^2|b, s_e^2] ≤ C/K.

All proofs are given in the appendix.

An immediate consequence of theorem 1 is that $\hat{s}_e^2$ is consistent.

Theorem 1 is nonstandard because the number of regressors increases with the sample size. In the special case that e_t is i.i.d. N(0, s_e^2) and the regressors are independent of the errors, the proof is straightforward. Under these special assumptions, in conventional matrix notation $\hat{s}_e^2 = y'[I-X(X'X)^{-1}X']y/(T-K) = e'[I-X(X'X)^{-1}X']e/(T-K)$. Because $I-X(X'X)^{-1}X'$ is idempotent with T-K degrees of freedom, $\hat{s}_e^2 / s_e^2$ is distributed $c_{T-K}^2/(T-K)$. Thus E[($\hat{s}_e^2$ - s_e^2)^2|b, s_e^2] = 2 $s_e^4/(T-K)$, so the result in theorem 1 holds with $C = 2r\, s_e^4/(1-r)$. The proof under the more general assumption 1 is, however, more involved.

The next theorem provides results for the OLS estimator and its forecast, $\hat{y}_{T+1|T}^{OLS}$.

Theorem 2 (OLS risk).

Under assumption 1, $R(\hat{b}, b) \to r\, s_e^2$, $r_G(\hat{b}, f_K) = r\, s_e^2$, and var($y_{T+1}$- $\hat{y}_{T+1|T}^{OLS}$)/ $s_e^2 \to 1+r$.

3.3 Results for Gaussian Empirical Bayes Estimators

The next two theorems establish the asymptotic Bayes risks of the Gaussian empirical Bayes estimators. Theorem 3 pertains to the parametric case, and theorem 4 pertains to the nonparametric case.

Theorem 3 (Parametric Empirical Bayes).

Suppose that the assumptions 1 – 5 and 7 hold. Then:

(i) $r_G(\hat{b}^{PEB}, f_K)$ - $r_G(\hat{b}^{NB}, f_K) \to 0$, and

(ii) $\inf_{\tilde{b}} \sup_{f_K} r_G(\tilde{b}, f_K) - r_G(\hat{b}^{NB}, f_K) \to 0$, where the supremum is taken over the class of likelihoods f_K that satisfy assumptions 1 – 4 with fixed constants.

Theorem 4 (Nonparametric Empirical Bayes).

Under assumptions 1-4, 6, and 7,

(i) $r_G(\hat{b}^{NSEB}, f_K)$ - $r_G(\hat{b}^{NB}, f_K) \to 0$, and

(ii) $\inf_{\tilde{b}} \sup_{f_K} r_G(\tilde{b}, f_K) - r_G(\hat{b}^{NB}, f_K) \to 0$, where the supremum is taken over the class of likelihoods f_K that satisfy assumptions 1 – 4 with fixed constants.

Part (i) of Theorem 3 states that the Bayes risk of the parametric EB estimator asymptotically equals the Gaussian Bayes risk of the infeasible estimator, $\hat{b}^{NB}$. By definition, $\hat{b}^{NB}$ is the Bayes estimator if G is known when the errors are normally distributed and are independent of the regressors. Thus the theorem implies that, if the errors are Gaussian and are independent of the regressors, the feasible estimator $\hat{b}^{PEB}$ is asymptotically optimal in the sense of Robbins (1964).

The theorem further states that the Bayes risk of the infeasible estimator $\hat{b}^{NB}$ is achieved even if the conditions for $\hat{b}^{NB}$ to be optimal (Gaussianity) are not met, as long as the assumptions of the theorem hold. Moreover, according to part (ii), this risk is achieved uniformly over distributions satisfying the stated assumptions. If $\{e_t\}$ has a nonnormal distribution, then the OLS estimators are no longer sufficient, and generally a lower Bayes risk can be achieved by using the Bayes estimator based on the true error distribution. Together these observations imply that $r_G(\hat{b}^{NB}, f_K)$ is an upper bound on the Bayes risk of the Bayes estimator under the prior G when $\{e_t\}$ is known but nonnormal. However, $\hat{b}^{PEB}$ is asymptotically optimal in the Gaussian case and its Bayes risk does not depend on the true error distribution asymptotically, so $\hat{b}^{PEB}$ is asymptotically minimax.

Because the Bayes risk function was derived from the forecasting problem, these statements about the properties of the parametric EB estimator carry over directly to forecasts based on the parametric EB estimator.

The interpretation of theorem 4 parallels that of theorem 3. In particular, $\hat{b}^{NSEB}$ is asymptotically optimal if the errors are normal and independent of the regressors. Asymptotically the Bayes risk of this estimator does not depend on the true distributions, as long as they satisfy the stated assumptions. Thus this estimator is asymptotically minimax for the family of distributions satisfying the assumptions of theorem 4.

Finally, because assumption 5 implies assumption 6, if the true distributions satisfy the conditions of theorem 3, the parametric and nonparametric EB estimators are asymptotically equivalent in the sense that they achieve the same asymptotic Bayes risk.

4. Frequentist Risk and Gaussian Equivariant Estimators

4.1 Frequentist Risk

As in the Bayesian case, the family of estimators considered in the frequentist formulation of the estimation problem is motivated by the Gaussian case. The sufficiency argument of section

2.4 applies here, so we consider estimators $\tilde{b}$ which are functions of the OLS estimator $\hat{b}$. Accordingly we write the frequentist estimation risk (2.5) as,

(4.1) $$R(b, \tilde{b}; f_K) = r K^{-1}\sum_{i=1}^{K} E[\tilde{b}_i(\hat{b}) - b_i]^2 .$$

where (4.1) differs from (2.5) by making explicit both the likelihood f_K under which the expectation is taken and the dependence of the estimator on $\hat{b}$.

4.2 Gaussian Equivariant Estimators

In the Gaussian case, the value of the likelihood f_K does not change under a simultaneous reordering of the cross sectional index i on $(\underline{X}_i, b_i)$, where $\underline{X}_i = (X_{i1},\ldots, X_{iT})$. More precisely, let **P** denote the permutation operator, so that $\mathbf{P}(\underline{X}_1, \underline{X}_2,\ldots, \underline{X}_K) = (\underline{X}_{i_1}, \underline{X}_{i_2}, \ldots, \underline{X}_{i_K})$, where $\{i_j, j=1,\ldots,K\}$ is a permutation of $\{1,\ldots,K\}$. The collection of all such permutations is a group, where the group operation is composition. The induced permutation of the parameters is $\mathbf{P}b$. In the Gaussian case, the likelihood constructed using $(\mathbf{P}X, \mathbf{P}b)$ equals the likelihood constructed using (X, b); that is, the likelihood is invariant to **P**.

Following the theory of equivariant estimation (e.g. Lehmann and Casella (1998, ch. 3)), this leads us to consider the set of estimators that are equivariant under any such permutation. An estimator $\tilde{b}(\hat{b})$ is equivariant under **P** if the permutation of the estimator constructed using $\hat{b}$ equals the (non-permuted) estimator constructed using the same permutation applied to $\hat{b}$. The set $\mathcal{B}$ of all estimators that are functions of $\hat{b}$ and are equivariant under the permutation group thus is, (4.2) $$\mathcal{B} = \{ \tilde{b}(\hat{b}) : \mathbf{P}[\tilde{b}(\hat{b})] = \tilde{b}(\mathbf{P}\hat{b}) \}.$$

An implication of equivariance is that the risk of an equivariant estimator is invariant under the permutation, that is,

(4.3) $$R(b, \tilde{b}; f_K) = R(\mathbf{P}b, \tilde{b}(\mathbf{P}\hat{b}); f_K)$$

for all **P**, cf. Lehmann and Casella (1998, Chapter 3, Theorem 2.7). Note that, in the problem at hand, because the risk is quadratic this invariance of the risk holds for all $\tilde{b} \in \mathcal{B}$ even if the motivating assumptions of the Gaussian case do not hold.

The set $\mathcal{B}$ contains the estimators commonly proposed for this problem, for example OLS, OLS with BIC model selection, ridge regression, and James-Stein estimation; $\hat{b}^{NSEB}$ is also in $\mathcal{B}$.

5. Results Concerning the Frequentist Risk

The next theorem characterizes the asymptotic limit of the frequentist risk of the minimum risk equivariant estimator. Let $\tilde{G}_K$ denote the (unknown) empirical cdf of the true coefficients $\{b_i\}$ for fixed K, and let $\hat{b}^{NB}_{\tilde{G}_K}$ denote the normal Bayes estimator constructed using (2.12), with the true empirical cdf $\tilde{G}_K$ replacing G. Also, let $\|x\|_2 = (x'x/K)^{1/2}$ for the K-vector x.

Theorem 5 (Minimum Risk Equivariant Estimator).

Suppose that the assumptions 1 – 4 and 7 hold. Then:

(i) $\inf_{\tilde{b}\in\mathcal{B}} R(b,\tilde{b};f_K) \geq R(b,\hat{b}^{NB}_{\tilde{G}_K};f_K) = r_{\tilde{G}_K}(\hat{b}^{NB}_{\tilde{G}_K},f)$ for all $b\in\mathbb{R}^K$ and for all K;

(ii) $\sup_{\|b\|_2\leq M} | R(b,\hat{b}^{NSEB};f_K) - \inf_{\tilde{b}\in\mathcal{B}} R(b,\tilde{b};f_K)| \rightarrow 0$ for all $M < \infty$; and

(iii) $\sup_{\|b\|_2\leq M}\{\sup_{f_K} | R(b,\hat{b}^{NSEB};f_K) - \inf_{\tilde{b}\in\mathcal{B}} R(b,\tilde{b};f_K)|\} \rightarrow 0$ for all $M < \infty$, where the supremum over f_K is taken over the class of likelihoods f_K which satisfy assumptions 1 – 4 with fixed constants.

Part (i) of this theorem provides a device for calculating a lower bound on the frequentist risk of any equivariant estimator in the Gaussian case. This lower bound can be expressed as the Bayes risk of the subjectivist normal Bayes estimator computed using the "prior" that equals the empirical cdf of the true coefficients; because this is computed using the true (unknown) empirical cdf, this is better thought of as a pseudo-Bayes risk. The estimator that achieves this in finite samples is the Bayes estimator constructed using the "prior" $\tilde{G}_K$, but because this prior is unknown this estimator is infeasible.

Part (ii) of the theorem shows that, in the Gaussian case, this optimal risk is achieved asymptotically by the nonparametric Gaussian simple empirical Bayes estimator. Moreover, this optimality holds uniformly for coefficient vectors in a normalized ball (of arbitrary radius) around the origin. Thus, in the Gaussian case the nonparametric Gaussian simple empirical Bayes estimator is asymptotically uniformly (over the ball) minimum risk equivariant.

Part (iii) of the theorem parallels the final parts of theorems 3 and 4, and shows that even outside the Gaussian case the frequentist risk of $\hat{b}^{NSEB}$ does not depend on f_K, as long as assumptions 1 – 4 hold. Because $\hat{b}^{NSEB}$ is optimal among equivariant estimators in the Gaussian case, and because its asymptotic risk does not depend on f_K, it is minimax among equivariant estimators.

6. Connecting the Frequentist and Bayesian Problems

The fact that $\hat{b}^{NSEB}$ is the optimal estimator in these two seemingly different estimation problems suggests that the problems themselves are related. It is well known that in conventional, fixed dimension parametric settings, by the Bernstein – von Mises argument, Bayes estimators and efficient frequentist estimators can be asymptotically equivalent. In these settings, a proper prior is dominated asymptotically by the likelihood. This is not, however, what is happening in this problem. Here, because the number of coefficients is increasing with the sample size and the coefficients are local to zero, the $\{b_i\}$ cannot be estimated consistently. Indeed, Stein's (1955) result that the OLS estimator is inadmissible holds here asymptotically, and the Bayes risks of the OLS and subjectivist Bayes estimators differ even asymptotically. Thus the standard argument, applicable to fixed parameter values, does not apply here.

Instead, the reason that these two problems are similar is that the frequentist risk for equivariant estimators is in effect the Bayes risk, evaluated at the empirical cdf $\tilde{G}_K$. For equivariant estimators, in the Gaussian case the i^{th} component of the frequentist risk (4.1), $E[\tilde{b}_i(\hat{b})-b_i]^2$ depends only on b_i. Thus we might write,

$$
\begin{aligned}
R(b,\tilde{b};f_K) &= r\,K^{-1}\sum_{i=1}^{K} E[\tilde{b}_i(\hat{b})-b_i]^2 \\
&= r\int E[\tilde{b}_1(\hat{b})-b_1]^2\, d\tilde{G}_K(b_1)
\end{aligned}
\tag{6.1}
$$

If the sequence of empirical cdfs { $\tilde{G}_K$ } has the weak limit G that is, $\tilde{G}_K \Rightarrow G$, and if the integrand in (6.1) is dominated, then

$$(6.2) \qquad R(b,\tilde{b};f_K) = r\int E[\tilde{b}(\hat{b})-b_1]^2\, d\tilde{G}_K(b_1) \Rightarrow r\int E[\tilde{b}(\hat{b})-b_1]^2\, dG(b_1)$$

which is the Bayes risk of $\tilde{b}$. This reasoning extends Edelman's (1988) argument linking the compound decision problem and the Bayes problem (for a narrow class of estimators) in the problem of estimating multiple means under a Gaussian likelihood.

This heuristic argument is made precise in the next theorem.

Theorem 6.

If $\tilde{G}_K \Rightarrow G$ and $\sup_K \|b\|_2 \le$ M, then $|\, R(b,\hat{b}^{NB}_{\tilde{G}_K};f_K) - r_G(\hat{b}^{NB},f)\,| \to 0$.

Thus, in the Gaussian case the frequentist risk of the subjectivist Bayes estimator $\hat{b}^{NB}_{\tilde{G}_K}$, based on the true empirical cdf $\tilde{G}_K$, and the Bayes risk of the subjectivist Bayes estimator $\hat{b}^{NB}$ based on its weak limit G, are the same asymptotically. It follows from theorems 3, 4 and 5 that this risk is also a lower bound on both the frequentist and Bayesian risks. This lower bound is achieved by the feasible nonparametric Gaussian simple empirical Bayes estimator, which, asymptotically, behaves as well as if the weak limit G were known.

7. Monte Carlo Analysis

7.1. Estimators

Parametric Gaussian EB estimator. The parametric Gaussian EB estimator examined in this Monte Carlo study is based on the parametric specification that $\{b_i\}$ are i.i.d. N(m, t^2). Using the normal approximating distribution for the likelihood, the marginal distribution of $\hat{b}_i$ is thus N($m, s_{\hat{b}}^2$), where $s_{\hat{b}}^2 = s_e^2 + t^2$. The parameters m and $s_{\hat{b}}^2$ are consistently estimated by $\hat{m} = K^{-1}\sum_{i=1}^{K}\hat{b}_i$ and $\hat{s}_{\hat{b}}^2 = (K-1)^{-1}\sum_{i=1}^{K}(\hat{b}_i - \hat{m})^2$. For the Monte Carlo analysis, we treat the sequence of constants s_K as a technical device and thus drop this term from (2.13). Accordingly, the parametric Gaussian empirical Bayes estimator, $\hat{b}^{PEB}$, is given by (2.14) with

(7.1) $$\hat{\ell}_K(\hat{b};\hat{q}) = -\frac{\hat{b} - \hat{m}}{\hat{s}_{\hat{b}}^2}.$$

Nonparametric Simple EB estimator. The nonparametric Gaussian simple EB estimator is computed as in (2.10) and (2.11), with some modifications. Following Härdle et. al. (1992), the score function is estimated using the bisquare kernel with bandwidth proportional to $(T/100)^{-2/7}$. Preliminary numerical investigation found advantages to shrinking the nonparametric score estimator towards the parametric Gaussian score estimator. We therefore use the modified score estimator,

(7.2) $$\hat{\ell}_{iK}^{s}(x) = l_T(x)\hat{\ell}_{iK}(x) + [1 - l_T(x)]\hat{\ell}_K(x;\hat{q})$$

where $\hat{\ell}_{iK}(x)$ is (2.18) implemented using the bisquare kernel nonparametric score estimator and s_K = 0, and, and $\hat{\ell}_K(x;\hat{q})$ is given in (7.1). The shrinkage weights are $l_T(x) = \exp[-½k^2(x-\hat{m})^2/s_{\hat{b}}^2$. Results are presented for various shrinkage parameters k; small values of k represent less shrinkage, and when k = 0, $\hat{\ell}_{iK}^s(x) = \hat{\ell}_{iK}(x)$.

Deconvolution EB estimator. An alternative to the nonparametric SEB estimator is to estimate the density g directly by nonparametric deconvolution, and then to use the estimated g in an empirical version of (2.10). Let $\hat{g}$ be such an estimator of g. The nonparametric deconvolution EB estimator is,

(7.3) $$\hat{b}_i^{NDEB} = \frac{\int b_i \hat{f}(\hat{b}_i - b_i)\hat{g}(b_i)\mathrm{d}b_i}{\int \hat{f}(\hat{b}_i - b_i)\hat{g}(b_i)\mathrm{d}b_i}.$$

Various approaches are available for estimating g. The specific deconvolution estimator considered here is constructed in the manner of Fan (1991) and Diggle and Hall (1993). If $\{\hat{b}_i\}$ are i.i.d. normal (conditional on b), then the marginal distribution of $\hat{b}_i$ is,

(7.4) $$m_f(x) = \int f(x-u)g(u)\mathrm{d}u.$$

Let $c_m(t) = \int m(x)\mathrm{e}^{-itx}\mathrm{d}x$, etc. Then (7.4) implies that $c_m(t) = c_f(t)c_g(t)$, so $c_g(t)$ = $c_m(t)/c_f(t)$. Let $\hat{m}$ be a kernel density estimator of m. This suggests the nonparametric estimator of the characteristic function of g, $\hat{c}_g(t) = c_{\hat{m}}(t)/c_{\hat{f}}(t)$. Following Diggle and Hall (1993), we therefore consider the nonparametric deconvolution estimator of g,

(7.5) $$\hat{g}(x) = \int w(t)\hat{c}_g(t)e^{itx}dx + \int[1-w(t)]c_{g*}(t)e^{itx}dx\ ,$$

where $w(t)$ is a weight function and g^* is a fixed density. Diggle and Hall(1991) choose $c_{g*}(t) = 0$ and $w(t) = \mathbf{1}(|t|\leq p_T)$, where $p_T > 0$ and $p_T \to \infty$ as $T \to \infty$.

No formal results are presented here for this estimator. If $\{\hat{b}_i\}$ are i.i.d., a result similar to Theorem 4 can be proven using the central limit theorem and arguments similar to those in Fan (1991) and Diggle and Hall (1993). However, extending this proof to the case that $\{\hat{b}_i\}$ are dependent appears to be difficult.

The nonparametric deconvolution EB estimator, $\hat{b}^{NDEB}$, is computed using (7.3) and (7.5), where the integrals are evaluated numerically. The kernel density estimator $\hat{m}$ was computed from $\{\hat{b}_i\}$ using a t-distribution kernel with five degrees of freedom and bandwidth $c(T/100)^{-2/7}/\hat{s}_{\hat{b}}$, where c is a constant (referred to below as the t-kernel bandwidth parameter). This heavy-tailed kernel was found to perform better than truncated kernels because $\hat{m}$ appears in the denominator of the EB estimate of the posterior mean. Diggle and Hall (1993) chose c_{g*} in (7.5) to be zero, so that the deconvolution estimator was shrunk towards a uniform distribution. However, numerical experimentation indicated that it is better to shrink towards the parametric Gaussian prior, so this is the choice of g^* used for the results here. The weight function $w(t)$ was chosen to be triangular so $w(0) = 1$ and $w(p_T) = 0$.

Both the nonparametric deconvolution EB and nonparametric simple EB estimators occasionally produced extremely large estimates, and some results were sensitive to these outliers. We therefore implemented the upper truncation $|\hat{b}_i^{EB}| \leq \max_i|\hat{b}_i|$ for all nonparametric estimators.

Other benchmark estimators. Results are also reported for some estimators that serve as benchmarks: the infeasible Bayes estimator, the OLS estimator, and the BIC estimator. The infeasible Bayes estimator is the Bayes estimator based on the true G and s_e^2; this is feasible only in a controlled experimental setting. The BIC estimator is the estimator that estimates b_i either by $\hat{b}_i$ or by zero, depending on whether this regressor is included in the regression according to the BIC criterion. Enumeration of all possible models and thus exhaustive BIC selection is possible in this design because of the orthonormality of the X's.

7.2 Experimental Design

The data were generated according to (1.1), with e_t i.i.d. N(0,1), where X_t are the K principal components of $\{W_t, t=1,...,T\}$, where W_{it} are i.i.d. N(0,1) and independent of $\{e_t\}$; X_t was rescaled to be orthonormal. The number of regressors was set at $K = rT$. Results are presented for $r = 0.4$ and $r = 0.7$.

Two sets of calculations were performed. The first examines the finite sample convergence of the Bayes risk of the various estimators to the Gaussian Bayes risk of the true Bayes estimator; that is, this calculation examines the relevance of theorems 3 and 4 to the finite sample behavior of these estimators. For these calculations, the parameters b_i were drawn from the mixture of normals distribution, (7.6) b_i i.i.d. $N(m_1, s_1^2)$ w.p. l and $N(m_2, s_2^2)$ w.p. $1-l$.

Six configurations of the parameters, taken from Marron and Wand (1992), were chosen to generate a wide range of distribution shapes. The densities are shown are shown in figure 1. The first sets $l = 1$, so that the b's are normally distributed. The second and third are symmetric and bimodal, and the fourth is skewed. The fifth density is heavy tailed, and the sixth is extremely so. In all of the experiments, the mean and variance parameters were scaled so that the population regression R^2 was 0.40.

The normal/mixed normal design allows analytic calculation of the Bayes risk for the (infeasible) Bayes estimator and the OLS estimator (where the risk only depends on the second moments). For the other estimators, the Bayes risk r_G was estimated by Monte Carlo simulation, with 1000 Monte Carlo repetitions, where each repetition entailed redrawing (b, X, e).

The second set of calculations evaluates the frequentist risk of the various estimators for a design in which the coefficients are fixed rather than drawn from a distribution. For these calculations, b_i was set according to

(7.7) $$b_i = \begin{cases} g, i = 1, \ldots, [lK] \\ 0, i = [lK]+1, \ldots, K \end{cases}$$

where l is a design parameter between 0 and 1, and g is chosen so that the population $R^2 = 0.4$. For these results, r was set at 0.4.

7.3 Results and Discussion

The Bayes estimation risk results are presented in table 1; results for r = 0.4 and r = 0.7 are shown in panels A and B, respectively. First consider the results in panel A. The Bayes risk of the OLS estimator is r = 0.4 for all sample sizes; the experimental design means that the asymptotic result in Theorem 2 holds exactly in finite samples. The Bayes estimators offer substantial improvements over OLS, with risks ranging from 0.25 to 0.12 (improvements of 40% – 70% relative to OLS). The BIC estimator generally performs worse than OLS, presumably because BIC is in part selecting variables that have small true coefficients but large estimated coefficients because of sampling error. The exception to this is when the b's are generated by the outlier distribution. Here 10% of the b's are drawn from a large-variance normal, and 90% of the b's are drawn from a small-variance normal concentrated around 0. Thus, most of the regression's predictive ability comes from a few regressors with large coefficients, and BIC does a relatively good job selecting these few regressors.

Two results stand out when looking at the performance of the empirical Bayes estimators. First, their performance is generally very close to the infeasible Bayes estimator for all of the sample sizes considered here, and thus they offer substantial improvement on the OLS and BIC estimators. The exception occurs when the b's are generated by the outlier distribution. In this case the empirical Bayes estimators achieve approximately half of the gain of the infeasible Bayes estimator, relative to OLS. For these outlier distributions, the BIC estimator dominates the empirical Bayes estimators. The second result that stands out is that the three empirical Bayes estimators have very similar performance. This is not surprising when the b's are generated from the Gaussian distribution, since in this case the parametric Gaussian empirical Bayes is predicated on the correct distribution. In this case, the similar performance of the non-parametric estimators suggests that little is lost ignoring this information. However, when the b's are generated by non-Gaussian distributions, the parametric Gaussian empirical Bayes estimator is misspecified. Yet, this estimator still performs essentially as well as the non-parametric estimators, and, except in the case of the outlier distribution, performs nearly as well as the optimal infeasible Bayes estimator.

The results in panel B, for which r = 0.7, present a similar picture. The Bayes risks of the OLS and BIC estimators is typically poor and is worse than the parametric or nonparametric EB estimators. The parametric and nonparametric EB estimators have Bayes risk approaching that of the true Bayes estimator except when the b's are generated by the outlier distribution.

The frequentist risk results are given in table 2. No prior is specified so the Bayes estimator is not relevant here. When l is small there are only a few non-zero (and large) coefficients, much like the b's generated by the outlier distribution. Thus, the results for l = 0.05 are much like those for the outlier distribution in table 1; BIC does well selecting the few non-zero coefficients; the empirical Bayes estimators perform well relative to OLS, but are dominated by BIC. However, the performance of BIC drops sharply as l increases; BIC and OLS are roughly comparable when l = .30, but when l = 0.50, the risk of BIC is 50% larger than the risk of OLS. In contrast, the empirical Bayes estimators work well for all values of l. For example, the (frequentist) risk of the nonparametric simple empirical Bayes estimator offer a 50% improvement on OLS when l is small, and more than a 75% improvement when l is large.

8. Application to Forecasting Monthly U.S. Macroeconomic Time Series

This section summarizes the results of using these methods to forecast monthly U.S. economic time series. The forecasts are based on the principal components of 151 macroeconomic time series. Forecasts based on the first few of these principal components from closely related data sets are studied in detail in Stock and Watson (1998, 1999). Here, we extend the analysis in those papers by considering forecasts based on all of the principal components.

8.1. Data

Forecasts were computed for four measures of aggregate real economic activity in the United States: total industrial production (ip); real personal income less transfers (gmyxpq); real manufacturing and trade sales (msmtq); and the number of employees on nonagricultural payrolls (lpnag). The forecasts were constructed using a set of 151 predictors that cover eight broad categories of available macroeconomic and financial time series. The series are listed in appendix B. The complete data set spans 1959:1-1998:12.

8.2. Construction of the Forecasts

Forecasts were constructed from regressions of the form

(8.1) $$y_{t+1} = b'X_t + e_{t+1},$$

where X_t is composed of the first K principal of the standardized predictors. The coefficient vector b was estimated by OLS and by the parametric and nonparametric simple empirical Bayes estimators. These estimators were implemented as in the Monte Carlo experiment. Results are presented for both one month ahead and one quarter ahead predictions. These latter results were calculated using quarterly aggregates of the data constructed using the final monthly observation of the quarter.

All forecasts are computed recursively (that is, in simulated real time) beginning in 1970:1. Thus, for example, to compute the forecasts for month T, principal components of the predictors were computed using data from 1960:1 through month T. The first $K = \min(151, rT)$ principal components were used as X_t, where $r = 0.4$. To capture serial correlation in the variables being predicted, residuals from univariate autoregressions were used for y_{t+1}. Thus, letting z_t denote the variable to be forecast, y_{t+1} was formed as the residual from the regression z_{t+1} onto $(1, z_t, z_{t-1},\ldots, z_{t-3})$ with data from t = 1960:1 through T-1. The regression coefficients in (5.1) were then estimated using the methods described above with data through from t = 1960:1 through T-1. These estimated coefficients, together with the coefficients from the autoregression were used to construct forecasts for z_{T+1}. This procedure was carried out for T = 1970:1 through the last available observation in 1998.

In addition, as a benchmark we report forecasts based on the first two principal components, a constant, and lags of z_t estimated by OLS, that is, OLS forecasts with $(X_{1t}, X_{2t}, 1, z_t,\ldots,z_{t-3})$ as regressors. Following Stock and Watson (1998), we refer to these as "DIAR" forecasts.

8.3 Results

Results are presented in table 3. The entries in this table are the mean square error of the simulated forecast errors relative to the mean square error from the univariate autoregression. Thus, for example, the first row of table 3 shows the results for the 1-month-ahead predictions of industrial production. The value of 1.01 under the column heading "OLS" means that the forecast constructed using the OLS estimates of b had a mean square error that was 1% greater than the forecasts that set $b = 0$ (the univariate autoregressive forecast). Results are also shown for the empirical Bayes estimators and for the DIAR estimator.

Several findings stand out in table 3. First, in all cases the empirical Bayes estimators improve upon OLS. Second, the relative MSE of the empirical Bayes estimators are always less than 1.0, so that these forecasts improve on the univariate autoregression. Third, as in the Monte Carlo experiment, the parametric and non-parametric empirical Bayes estimators have nearly identical performance. Finally, the DIAR models yielded the most accurate forecasts. Apparently, it is better to forecast using only the first two principal components of the predictors with no shrinkage, than to use many of the principal components and shrink them toward a common value.

Taken as a whole the table suggests only modest improvement of the empirical Bayes estimators relative to the univariate autoregression. This is somewhat surprising given the performance of the Empirical Bayes estimators in the Monte Carlo experiments reported in section 4.3. The explanation seems to be that the predictive power of the regression (measured by the regression R^2) is not as great in as in the Monte Carlo design. In the Monte Carlo experiment, the R^2 was 0.40, and for the series considered here it is considerably less than 0.40. For example, suppose for a moment that the DIAR results give a good estimate of the forecastibility of y given all of the predictors. Thus, for the 1-month ahead forecasts, the R^2 is approximately 10% – 15%. A calculation shows that if the population R^2 is 15% and r = 0.4, then the asymptotic relative efficiency of the empirical Bayes estimators is only 0.95, and deteriorates to 0.98 when R^2 falls to 10%.

The final question addressed in this section is whether the empirical Bayes methods can be used to improve upon the DIAR models. To answer this question the forecasting experiment was repeated, but now using the DIAR model as baseline regression rather than the univariate autoregression. Thus, residuals from the DIAR forecasts were used for y_{t+1} in the empirical Bayes regressions. The results for this experiment are shown in table 4. There is some evidence that the EB estimators can yield small improvements on the DIAR model. For example, PEB yields an average 3% improvement over DIAR.

9. Discussion and Conclusion

This paper studied the problem of prediction in a linear regression model with a large number of predictors. This framework leads to a natural integration of frequentist and Bayesian methods. In particular, in the Gaussian case, the limiting frequentist risk of permutation-equivariant estimators and the limiting Bayes risk share a lower bound which is the risk of the subjectivist Bayes estimator constructed using a “prior” that equals the limiting empirical distribution of the true regression coefficients. This bound is achieved by the empirical Bayes estimators laid out in this paper. The empirical Bayes estimators use the large number of estimated regression coefficients to estimate this "prior." These results differ in an important way from the usual asymptotic analysis of Bayes estimators in finite dimensional settings, in which the likelihood dominates the prior distribution. Here the number of parameters grows proportionally to the sample size so that the prior affects the posterior, even asymptotically.

The Monte Carlo analysis suggested that the proposed empirical Bayes methods work well in finite samples for a range of distributions of the regression coefficients. An important exception was a distribution that generated a very few large non-zero coefficients with the remaining coefficients very close to zero. Only in this case did choosing the regressors by BIC perform better than the empirical Bayes estimators.

Although macroeconomic forecasting motivated our interest in the methods developed here, the theoretical results also contribute to the econometric literature on regression with many unknown parameters. Thus, for example, these methods may prove useful for instrumental variable models with many instruments (e.g., Angrist and Krueger (1991), Bekker (1994), Chamberlain and Imbens (1996)).

There are several unfinished extensions of this work. The theoretical analysis relied on martingale difference regression errors and orthonormal regressors. The assumption of martingale difference errors prevents these results from applying directly to multiperiod forecasting, a problem of practical interest. Within the framework of orthonormal regressors, one might want to model the potential forecasting importance of the factors as diminishing with their contribution to the R^2 of the original data. It is straightforward to do this using parametric empirical Bayes techniques, but it is less clear how to extend this idea to the nonparametric empirical Bayes or equivariant estimation problems. Similarly, although the assumption of orthonormal regressors coincides with the factor structure used in the empirical application, in other applications it might be more natural to forecast using the original, nonorthogonalized regressors.

Finally, the empirical Bayes estimators yielded considerable improvement in the Monte Carlo design – indeed they approached the efficiency of the infeasible "true" Bayes estimator – yet they delivered only small improvements in the empirical application. This suggests that the empirical finding is not the result of using an inefficient forecast, but rather that there simply is little predictive content in these macroeconomic principal components beyond the first few. If true, this has striking and, we believe, significant implications for empirical macroeconomics and large-model forecasting. Additional analysis remains, however, before we can be confident of this intriguing negative finding.

References

Angrist, J. and A. Krueger (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106, 979-1014.

Bekker, P.A. (1994), "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657-681.

Berry, A.C. (1941), "The accuracy of the Gaussian approximation to the sum of independent variates", *Transactions of the American Mathematical Society* 49, 122-136.

Bickel, P. (1982), "On Adaptive Estimation," *Annals of Statistics*, 10, 647-671.

Bickel, P., C.A.J. Klaassen, Y. Ritov, and J.A. Wellner (1993), *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD: Johns Hopkins University Press.

Billingsley, P. (1968), *Convergence of Probability Measures.* New York: Wiley.

Billingsley, P. (1995), *Probability and Measure, third edition*. New York: Wiley.

Birkel, T. (1988), "On the convergence rate in the central limit theorem for associated processes," *Annals of Probability* 16, 1685-1698.

Bolthausen, E. (1980), "The Berry-Esseen theorem for functionals of discrete Markov chains," *Z. Wahrsch. verw. Gebiete* 54, 59-73.

Bolthausen, E. (1982a), "Exact convergence rates in some martingale central limit theorems," *Annals of Probability* 10, 672-688.

Bolthausen, E. (1982b), "The Berry-Esseen theorem for strongly mixing Harris recurrent Markov chains," *Z. Wahrsch. verw. Gebiete* 60, 283-289.

Carlin, B.P. and T.A. Louis (1996), *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton, FL: Chapman and Hall.

Chamberlain, G. and G.W. Imbens (1996), "Hierarchical Bayes Models with Many Instrumental Variables," NBER Technical Working Paper 204.

Cramer, H. (1937), *Random Variables and Probability Distributions.* Cambridge, U.K.: Cambridge Tracts.

Diggle, P.J. and P. Hall (1993), "A Fourier Approach to Nonparametric Deconvolution of a Density Estimate," *Journal of the Royal Statistical Society, Series B* 55, 523-531.

Doukhan, P. (1994), *Mixing: Properties and Examples*. New York: Springer-Verlag.

Dudley, R. (1999), *Uniform Central Limit Theory*. New York: Cambridge University Press.

Edelman, D. (1988), "Estimation of the Mixing Distribution for a Normal Mean with Applications to the Compound Decision Problem," *Annals of Statistics* 16, 1609-1622.

Efron, B. and C.N. Morris (1972), "Empirical Bayes Estimators on Vector Observations - An Extension of Stein's Method," *Biometrika*, 59, 335-347.

Esseen, C.-G. (1945), "Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law," *Acta Mathematica* 77, 1-125.

Fan, J. (1991), "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems", *Annals of Statistics* 19, 1257-1272

Feller, W. (1971), *An Introduction to Probability Theory and its Applications, Volume 2 (Second Edition)*. New York: Wiley.

George, E.I. (1999), "Comment on *Bayesian Model Averaging*," *Statistical Science* 14, no. 382, 409-412.

George, E.I. and D.P. Foster (2000), "Calibration and Empirical Bayes Variable Selection," manuscript, University of Texas – Austin.

Götze, F. (1991), "On the rate of convergence in the multivariate CLT," *Annals of Probability*, 724-739.

Götze, F. and C. Hipp (1983), "Asymptotic expansions for sums of weakly dependent random vectors," *Z. Wahrsch. verw. Gebiete* 64, 211-239.

Hall, P. and C.C. Heyde (1980), *Martingale Limit Theory and its Application*. New York: Academic Press.

Härdle, W., J. Hart, J.S. Marron, and A.B. Tsybakov (1992), "Bandwidth Choice for Average Derivative Estimation," *Journal of the American Statistical Association* 87, 218-226.

Hoeting, J.A., D. Madigan, A.E. Raftery, and C.T. Volinsky (1999), "Bayesian Model Averaging: A Tutorial," *Statistical Science* 14, no. 38, 382-401.

James, W. and C. Stein (1960), "Estimation with Quadratic Loss," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 361-379.

Lehmann, E.L. and G. Casella (1998), *Theory of Point Estimation, Second Edition*. New York: Springer-Verlag.

Maritz, J.S. and T. Lwin (1989), *Empirical Bayes Methods, Second Edition*. London: Chapman and Hall.

Marron, J.S. and M.P. Wand (1992), Exact Mean Integrated Squared Error," *The Annals of Statistics*, 20, 712-736.

Philipp, W. (1969), "The remainder in the central limit theorem for mixing stochastic processes," *Annals of Mathematical Statistics* 40, 601-609.

Rio, E. (1996), "Sur le theoreme de Berry-Esseen pour les suites faiblement dependantes," *Probability Theory and Related Fields* 104, 255-282.

Robbins, H. (1955), "An Empirical Bayes Approach to Statistics," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, 157-164.

Robbins, H. (1964), "The Empirical Bayes Approach to Statistical Problems," *Annals of Mathematical Statistics*, 35, 1-20.

Stein, C. (1955), "Inadmissibility of the Usual Estimator for the Mean of Multivariate Normal Distribution," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 197-206.

Stock, J.H. and M.W. Watson (1998), "Diffusion Indexes," manuscript.

Stock, J.H. and M.W. Watson (1999), "Forecasting Inflation," *Journal of Monetary Economics* 44, 293-335.

Tikhomirov, A. (1980), "On the convergence rate in the central limit theorem for weakly dependent random variables," *Theoretical Probability and its Applications* 25, 790-809.

Van der Vaart, A.W. (1988), *Statistical Estimation in Large Parameter Spaces*. Stichting Mathematisch Centrum, Amsterdam.

Table 1
Bayes Estimation Risk of Various Estimators
Regression $R^2 = 0.40$, $s_e^2 = 1.0$

A. $r = 0.4$

β Distribution	Estimators					
	Bayes	OLS	BIC	PEB	NSEB	NDEB
i. T=50						
Gaussian	0.25	0.40	0.54	0.29	0.29	0.31
Bimodal	0.24	0.40	0.62	0.29	0.28	0.29
Separated Bimodal	0.22	0.40	0.69	0.29	0.28	0.28
Asymmetric Bimodal	0.24	0.40	0.61	0.28	0.28	0.29
Kurtotic	0.24	0.40	0.44	0.28	0.28	0.32
Outlier	0.13	0.40	0.17	0.25	0.24	0.28
ii. T=100						
Gaussian	0.25	0.40	0.56	0.27	0.27	0.28
Bimodal	0.24	0.40	0.63	0.27	0.27	0.27
Separated Bimodal	0.22	0.40	0.69	0.27	0.26	0.25
Asymmetric Bimodal	0.23	0.40	0.62	0.26	0.26	0.26
Kurtotic	0.24	0.40	0.46	0.26	0.27	0.29
Outlier	0.12	0.40	0.16	0.24	0.23	0.27
iii. T=200						
Gaussian	0.25	0.40	0.57	0.26	0.26	0.26
Bimodal	0.24	0.40	0.64	0.26	0.26	0.26
Separated Bimodal	0.22	0.40	0.70	0.26	0.25	0.24
Asymmetric Bimodal	0.23	0.40	0.63	0.25	0.25	0.25
Kurtotic	0.24	0.40	0.49	0.26	0.26	0.27
Outlier	0.12	0.40	0.16	0.24	0.22	0.27
iv. T=400						
Gaussian	0.25	0.40	0.59	0.25	0.26	0.26
Bimodal	0.24	0.40	0.65	0.26	0.26	0.25
Separated Bimodal	0.22	0.40	0.69	0.25	0.24	0.25
Asymmetric Bimodal	0.23	0.40	0.65	0.24	0.25	0.24
Kurtotic	0.24	0.40	0.51	0.25	0.25	0.26
Outlier	0.12	0.40	0.17	0.25	0.21	0.30

Table 1 Continued

A. r = 0.7

β Distribution	Estimators					
	Bayes	OLS	BIC	PEB	NSEB	NDEB
i. T=50						
Gaussian	0.34	0.70	0.69	0.42	0.43	0.39
Bimodal	0.33	0.70	0.75	0.43	0.43	0.39
Separated Bimodal	0.32	0.70	0.81	0.43	0.42	0.38
Asymmetric Bimodal	0.32	0.70	0.73	0.41	0.41	0.36
Kurtotic	0.34	0.70	0.61	0.42	0.42	0.39
Outlier	0.19	0.70	0.34	0.39	0.38	0.34
ii. T=100						
Gaussian	0.34	0.70	0.68	0.38	0.39	0.37
Bimodal	0.34	0.70	0.75	0.38	0.38	0.36
Separated Bimodal	0.33	0.70	0.81	0.39	0.38	0.36
Asymmetric Bimodal	0.32	0.70	0.74	0.36	0.37	0.35
Kurtotic	0.33	0.70	0.60	0.38	0.39	0.37
Outlier	0.19	0.70	0.32	0.36	0.35	0.33
iii. T=200						
Gaussian	0.34	0.70	0.68	0.36	0.37	0.36
Bimodal	0.34	0.70	0.75	0.36	0.37	0.36
Separated Bimodal	0.32	0.70	0.80	0.36	0.36	0.35
Asymmetric Bimodal	0.32	0.70	0.73	0.34	0.35	0.34
Kurtotic	0.33	0.70	0.59	0.36	0.37	0.36
Outlier	0.19	0.70	0.29	0.35	0.33	0.33
iv. T=400						
Gaussian	0.34	0.70	0.68	0.35	0.36	0.35
Bimodal	0.34	0.70	0.74	0.35	0.36	0.35
Separated Bimodal	0.33	0.70	0.79	0.35	0.35	0.35
Asymmetric Bimodal	0.32	0.70	0.74	0.33	0.34	0.33
Kurtotic	0.33	0.70	0.60	0.35	0.36	0.35
Outlier	0.18	0.70	0.27	0.35	0.32	0.33

Notes: The values shown in the table are the Bayes risk $r_G(\tilde{b}, f_K)$ where the distribution of the coefficients is shown in first column. The estimators are the exact (infeasible) Bayes estimator, OLS, BIC models selection over all possible regressions, the parametric Gaussian simple empirical Bayes estimator (PEB), the nonparametric Gaussian simple empirical Bayes estimator (NSEB), and the nonparametric deconvolution Gaussian empirical Bayes estimator (NDEB).

Table 2
Classical Estimation Risk of Various Estimators
Regression R^2 = 0.40, s_e^2 = 1.0, T = 200, r = 0.40

$$b_i = g \times 1(i \le l K)$$

λ	Estimator				
	OLS	BIC	PEB	NSEB	NDEB
0.05	0.40	0.08	0.26	0.20	0.27
0.10	0.40	0.11	0.25	0.19	0.29
0.20	0.40	0.28	0.24	0.21	0.27
0.30	0.40	0.42	0.23	0.21	0.24
0.40	0.40	0.52	0.21	0.21	0.22
0.50	0.40	0.58	0.20	0.19	0.20
0.60	0.40	0.63	0.17	0.18	0.17
0.70	0.40	0.66	0.15	0.15	0.15
0.80	0.40	0.69	0.12	0.12	0.11
0.90	0.40	0.71	0.08	0.09	0.08

Notes: The values shown in the table are the classical risk, $R(\tilde{b}, b)$, where the first lK values of b take on the value g and the remaining values are 0. The estimators are OLS, BIC models selection over all possible regressions, the parametric Gaussian simple empirical Bayes estimator (PEB), the nonparametric simple Gaussian empirical Bayes estimator (NSEB), and the nonparametric deconvolution Gaussian empirical Bayes estimator (NDEB).

Table 3

Simulated Out-of-Sample Forecasting Results with r = 0.4

Mean Square Errors Relative to Univariate Autoregression

Series	Forecast Method			
	OLS	PEB	NSEB	DIAR
1 Month Ahead Forecasts				
Industrial Production	1.01	0.94	0.94	0.89
Personal Income	1.07	0.98	0.98	0.91
Mfg. & Trade Sales	1.03	0.94	0.94	0.88
Nonag. Employment	1.04	0.95	0.95	0.82
1 Quarter Ahead Forecasts				
Industrial Production	0.95	0.94	0.94	0.72
Personal Income	0.96	0.95	0.95	0.81
Mfg. & Trade Sales	0.92	0.91	0.91	0.72
Nonag. Employment	1.02	0.96	0.96	0.73

Note: The table entries show the simulated out-of-sample forecast mean square error relative the mean square forecast error for a univariate autoregression. All forecasts were computed using recursive methods described in the text with a sample period beginning in 1960:1. The simulated out-of-sample forecast period is 1970:1-1998:12.

Table 4

Simulated Out-of-Sample Forecasting Results with r = 0.4

Mean Square Errors Relative to DIAR forecasts

Series	Forecast Method		
	OLS	PEB	NSEB
1 Month Ahead Forecasts			
Industrial Production	1.04	0.95	0.97
Personal Income	1.11	0.99	1.00
Mfg. & Trade Sales	1.03	0.96	0.96
Nonag. Employment	1.11	0.98	0.98
1 Quarter Ahead Forecasts			
Industrial Production	1.06	0.95	0.95
Personal Income	1.02	0.98	1.00
Mfg. & Trade Sales	0.98	0.99	0.98
Nonag. Employment	1.19	1.01	1.01

Note: The table entries show the simulated out-of-sample forecast mean square error relative the mean square forecast error for the DIAR model with 2 factors. All forecasts were computed using recursive methods described in the text with a sample period beginning in 1960:1. The simulated out-of-sample forecast period is 1970:1-1998:12.

Figure 1. Distribution of Regression Coefficients