EMPIRICAL BAYES FORECASTS OF ONE TIME SERIES USING MANY PREDICTORS

Thomas Knox
James H. Stock
Mark W. Watson

Empirical Bayes Forecasts of One Time Series Using Many Predictors
Thomas Knox, James H. Stock and Mark W. Watson
NBER Technical Working Paper No. 269
March 2001
JEL No. C32, E37

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.

Thomas Knox
Department of Economics
Harvard University

James H. Stock
Kennedy School of Government
Harvard University
and NBER

Mark W. Watson
Woodrow Wilson School
Princeton University
and NBER

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 $\text{tr}(V_{\tilde{b}})$, 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 $\text{tr}(V_{\tilde{b}})$. 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 $\text{tr}(V_{\tilde{b}})$

among all equivariant estimators is the minimum risk equivariant estimator. If this estimator

---

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, $\int \text{tr}(V_{\tilde{b}})dG_b(\boldsymbol{b}_1)\cdots dG_b(\boldsymbol{b}_K)$, rather than the frequentist risk $\text{tr}(V_{\tilde{b}})$. 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 $\boldsymbol{b}$. 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 $T \rightarrow \infty$, 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, $K/T \rightarrow r$ as $T \rightarrow \infty$. 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) $\quad\quad y_{t+1} = \boldsymbol{b}'X_t + \boldsymbol{e}_{t+1},$

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

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

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

The forecast risk is the expected loss,

(2.3) $\quad\quad EL(\tilde{\boldsymbol{b}},\boldsymbol{b}) = \boldsymbol{s}_e^2 + E(\tilde{\boldsymbol{b}} - \boldsymbol{b})'H_T(\tilde{\boldsymbol{b}} - \boldsymbol{b}),$

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

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

6

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 $\text{tr}(V_{\tilde{b}})$, 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,

$$(2.4) \qquad b = b/\sqrt{T} \text{ and } \tilde{b} = \tilde{b}/\sqrt{T} .$$

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

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

where b_i is the ith element of b, etc.

2.2 Asymptotic Nesting

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

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

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

Under this nesting, if each $\boldsymbol{b}_i$ 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 $E(\hat{b}-b)(\hat{b}-b)' = \boldsymbol{s}_e^2 I_K$. Also let $\hat{\boldsymbol{s}}_e^2$ be the usual estimator of $\boldsymbol{s}_e^2$,

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

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 $\boldsymbol{f}_K(z) = \prod_{i=1}^{K} \boldsymbol{f}(z_i)$, where $\boldsymbol{f}$ is the normal density with mean zero and variance s_e^2. Thus, in the Gaussian case $f_K = \boldsymbol{f}_K$. Note that, if K were fixed, standard central limit theory would imply that $\boldsymbol{f}_K$ 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)
$$r_G(\tilde{b}, f_K) = r \int K^{-1} \sum_{i=1}^{K} E(\tilde{b}_i - b_i)^2 \, dG_K(b)$$
$$= r \int K^{-1} \sum_{i=1}^{K} \int (\tilde{b}_i(\hat{b}) - b_i)^2 f_K(\hat{b} - b) \, d\hat{b} dG_K(b)$$

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,

$$(2.10) \qquad \frac{\int x \boldsymbol{f}_K(\hat{b}-x)\mathrm{d}G_K(x)}{\int \boldsymbol{f}_K(\hat{b}-x)\mathrm{d}G_K(x)} .$$

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,

$$(2.11) \qquad m_f(x) = \int \boldsymbol{f}_K(x-b) dG_K(b) ,$$

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,

$$(2.12) \qquad \hat{b}^{NB} = \hat{b} + \boldsymbol{s}_e^2 \ell_f(\hat{b}) .$$

10

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 $\boldsymbol{s}_e^2$. Thus the score ℓ_f is unknown, and the estimator (2.12) is infeasible. However, both the score and $\boldsymbol{s}_e^2$ can be estimated. Moreover, although the derivation of (2.12) relies on $f_K = f_K$, as was mentioned $\boldsymbol{f}_K$ might be a plausible large sample approximation to f_K. This suggests using an estimator of the form (2.12), with ℓ_f and $\boldsymbol{s}_e^2$ 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, $\boldsymbol{s}_e^2$ is estimated by $\hat{\boldsymbol{s}}_e^2$, 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 $\boldsymbol{f}_K$ to the likelihood, this in turn provides a parametric approximation to the marginal distribution of $\hat{b}$,

$m_f(x;\theta) = \int \boldsymbol{f}_K(x-b)dG_K(b;\boldsymbol{q})$. The parametric Gaussian empirical Bayes estimator is computed by substituting estimates of $\boldsymbol{s}_e^2$ and $\boldsymbol{q}$ 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,

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

where $m_{\hat{f}}(x;\hat{\boldsymbol{q}}) = \int \hat{\boldsymbol{f}}_K(x-b)\,dG_K(b;\hat{\boldsymbol{q}})$ and $m'_{\hat{f}}(x;\hat{\boldsymbol{q}}) = \int \hat{\boldsymbol{f}}'_K(x-b)\,dG_K(b;\hat{\boldsymbol{q}})$, where $\hat{\boldsymbol{f}}_K$ denotes $\boldsymbol{f}_K$

with $\boldsymbol{s}_e^2$ estimated, that is, $\hat{\boldsymbol{f}}_K(u) = (2\boldsymbol{p}\hat{\boldsymbol{s}}_e^2)^{-K/2}\exp(-u'u/2\hat{\boldsymbol{s}}_e^2)$, 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 $\hat{\boldsymbol{s}}_e^2$; thus,

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

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 $w'(z) = dw(z)/dz$, and let h_K denote the kernel bandwidth sequence.

Define

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

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

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

The nonparametric score estimator considered here is,

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

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,

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

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

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 $\{Z_s, s=a, ..., b\}$.

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

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

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, $| p_{itK}^X(x)| \leq C$, $| p_{ijtK}^X(x_i \mid x_j)| \leq C$

for $i \neq j$, and $\int| \dfrac{d^2}{de^2} p_{itK}^e(e)| \, de \leq C$.

The next assumption restricts the cross sectional dependence among $\{X_{it}\}$ using a conditional maximal correlation coefficient condition. Let $\underline{X}_i = (X_{i1},..., X_{iT})$ and let $\mathcal{F}_a^b$ be the σ-field generated by the random variables $\{\underline{X}_i, i = a,...,b\}$, and define the cross sectional conditional maximal correlation coefficients $\{t_n\}$ to be a sequence satisfying $\sup_m \sup_{y \in \mathcal{F}_1^m, x \in \mathcal{F}_{m+n}^\infty} |\mathrm{corr}(x,y|\underline{X}_j)| \leq t_n$ 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 $\text{var}(b_i) = \boldsymbol{s}_b^2 < \infty$.

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

(iii) G has density $g(b; \boldsymbol{q})$ which is Lipschitz continuous in $\boldsymbol{q}$ uniformly over b and $\boldsymbol{q}$, i.e.

$\sup_{b,\boldsymbol{q}\in\Theta}\sup_{\boldsymbol{q}'\in\Theta}|g(b;\boldsymbol{q}) - g(b;\boldsymbol{q}')| \leq C\|\boldsymbol{q}-\boldsymbol{q}'\|$.

(iv) There exists an estimator $\hat{\boldsymbol{q}} = \boldsymbol{q}(\hat{b},\hat{\boldsymbol{s}}_e^2)$ such that, for all K sufficiently large,

$\mathrm{E}[K\|\hat{\boldsymbol{q}}-\boldsymbol{q}\|^2] \leq C < \infty$, where the expectation is taken either under (f_K, G) or under $(\boldsymbol{f}_K, G)$.

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

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, $\hat{\boldsymbol{s}}_e^2$ is consistent.

Theorem 1 (standard error of the regression).

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

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 $E[(\hat{s}_e^2 - s_e^2)^2|b,s_e^2] = 2s_e^4/(T-K)$, so the result in theorem 1 holds with $C = 2rs_e^4/(1-r)$. 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, $R(\hat{b},b) \rightarrow rs_e^2$, $r_G(\hat{b},f_K) = rs_e^2$, and $var(y_{T+1}- \hat{y}_{T+1|T}^{OLS})/s_e^2 \rightarrow 1+r$.

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.

17

Theorem 3 (Parametric Empirical Bayes).

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

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

(ii) $\inf_{\tilde{b}} \sup_{f_K} r_G(\tilde{b}, f_K) - r_G(\hat{b}^{NB}, \mathbf{f}_K) \to 0$, 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) $r_G(\hat{b}^{NSEB}, f_K) - r_G(\hat{b}^{NB}, \mathbf{f}_K) \to 0$, and

(ii) $\inf_{\tilde{b}} \sup_{f_K} r_G(\tilde{b}, f_K) - r_G(\hat{b}^{NB}, \mathbf{f}_K) \to 0$, 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

18

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 $r_G(\hat{b}^{NB}, \boldsymbol{f}_K)$ is an upper bound on the Bayes risk of the Bayes estimator under the prior G when $\{\boldsymbol{e}_t\}$ 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,

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

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 $\mathbf{f}_K$ does not change under a simultaneous

reordering of the cross sectional index i on $(\underline{X}_i, b_i)$, where $\underline{X}_i = (X_{i1},..., X_{iT})$.   More precisely, let $\mathbf{P}$

denote the permutation operator, so that $\mathbf{P}(\underline{X}_1, \underline{X}_2,..., \underline{X}_K) = (\underline{X}_{i_1}, \underline{X}_{i_2},..., \underline{X}_{i_K})$, where $\{i_j, j=1,...,K\}$ is

a permutation of $\{1,...,K\}$. 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 $\mathbf{P}$.

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

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

(iii) $\sup_{\|b\|_2 \leq M} \{ \sup_{f_K} | R(b, \hat{b}^{NSEB}; f_K) - \inf_{\tilde{b} \in \mathcal{B}} R(b, \tilde{b}; \boldsymbol{f}_K) | \} \rightarrow 0$ 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 ith 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,

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

23

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

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

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 $|R(b,\hat{b}_{\tilde{G}_K}^{NB};\boldsymbol{f}_K) - r_G(\hat{b}^{NB},\boldsymbol{f})| \to 0$.

Thus, in the Gaussian case the frequentist risk of the subjectivist Bayes estimator $\hat{b}_{\tilde{G}_K}^{NB}$, 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. $N(\boldsymbol{m}, \boldsymbol{t}^2)$. Using the normal approximating distribution for the likelihood, the marginal distribution of $\hat{b}_i$ is thus $N(\boldsymbol{m}, \boldsymbol{s}_{\hat{b}}^2)$, where $\boldsymbol{s}_{\hat{b}}^2 = \boldsymbol{s}_e^2 + \boldsymbol{t}^2$. The parameters $\boldsymbol{m}$ and $\boldsymbol{s}_{\hat{b}}^2$ are consistently estimated by

$\hat{\boldsymbol{m}} = K^{-1} \sum_{i=1}^{K} \hat{b}_i$ and $\hat{\boldsymbol{s}}_{\hat{b}}^2 = (K-1)^{-1} \sum_{i=1}^{K} (\hat{b}_i - \hat{\boldsymbol{m}})^2$. 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)
$$\hat{\ell}_K(\hat{b}; \hat{\boldsymbol{q}}) = -\frac{\hat{b} - \hat{\boldsymbol{m}}}{\hat{\boldsymbol{s}}_{\hat{b}}^2}.$$

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

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 $l_T(x) = \exp[-\frac{1}{2}k^2(x-\hat{m})^2/s_{\hat{b}}^2]$.

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,

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

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,

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

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,

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

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 $\boldsymbol{s}_e^2$; 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 $\boldsymbol{e}_t$ 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 $\{\boldsymbol{e}_t\}$; X_t was rescaled to be orthonormal. The number of regressors was set at $K = \boldsymbol{r}T$. Results are presented for $\boldsymbol{r} = 0.4$ and $\boldsymbol{r} = 0.7$.

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 $\boldsymbol{b}_i$ were drawn from the mixture of normals distribution,

(7.6) $\qquad \boldsymbol{b}_i$ i.i.d. $N(\boldsymbol{m_1}, \boldsymbol{s}_1^2)$ w.p. $\boldsymbol{l}$ and $N(\boldsymbol{m_2}, \boldsymbol{s}_2^2)$ w.p. 1-$\boldsymbol{l}$.

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 $\boldsymbol{l} = 1$, so that the $\boldsymbol{b}$'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 ($\boldsymbol{b}$, X, $\boldsymbol{e}$).

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, $\boldsymbol{b}_i$ was set according to

(7.7) $\qquad \boldsymbol{b}_i = \begin{cases} \boldsymbol{g}, i = 1,...,[\boldsymbol{l} K] \\ 0, i = [\boldsymbol{l} K]+1,..., K \end{cases}$

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

31

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 $K = \min(151, \boldsymbol{r}T)$ principal components were used as X_t, where $\boldsymbol{r} = 0.4$. 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 $\boldsymbol{b}$ had a mean square error that was 1% greater than the forecasts that set $\boldsymbol{b} = 0$ (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

35

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, $r = 0.40$

$$\boldsymbol{b}_i = \boldsymbol{g} \times 1(i \leq \boldsymbol{l}\,K)$$

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, $R(\tilde{\boldsymbol{b}}, \boldsymbol{b})$, where the first $\boldsymbol{l}\,K$ values of $\boldsymbol{b}$ take on the value $\boldsymbol{g}$ 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

a. Gaussian
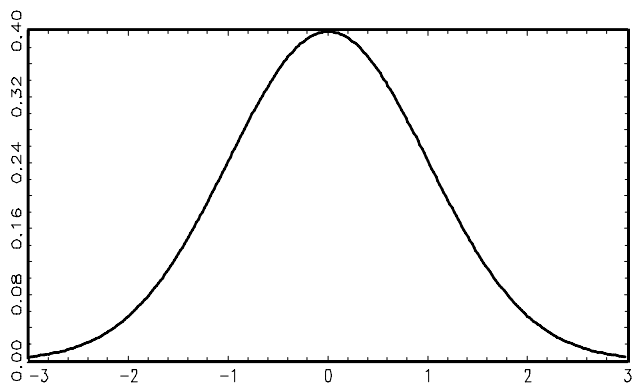
b. Bimodal
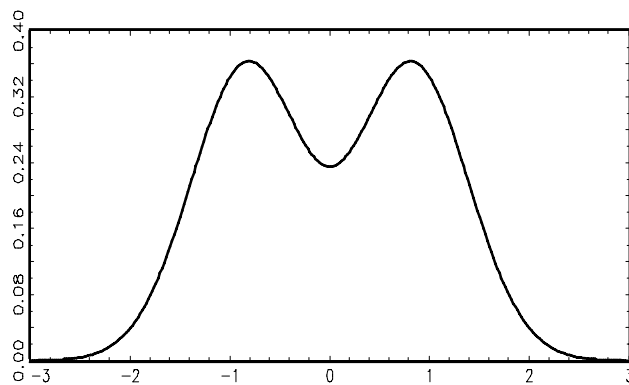
c. Separated Bimodal
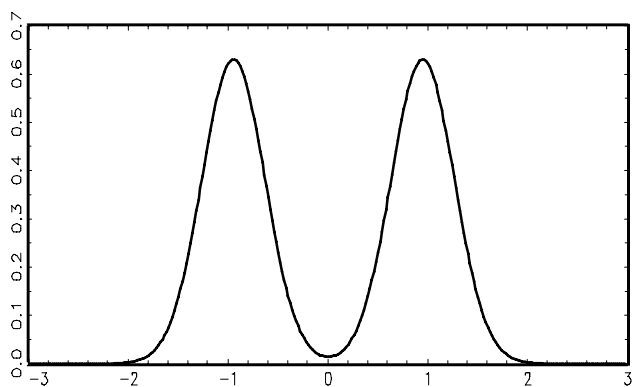
d. Asymmetric Bimodal
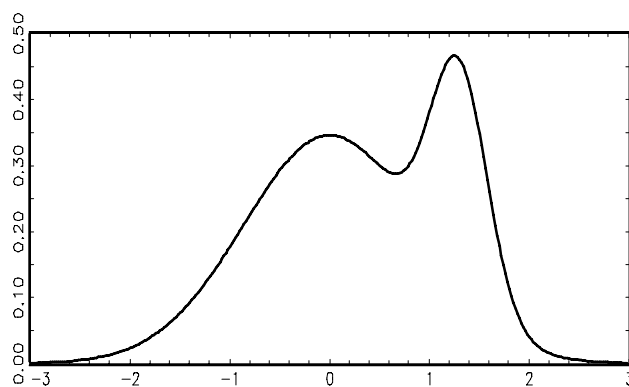
e. Kurtotic
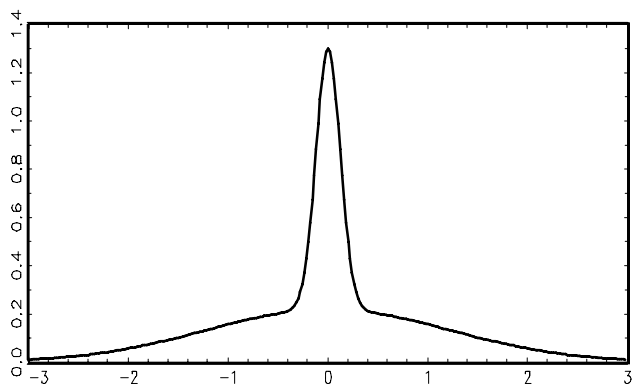
f. Outlier
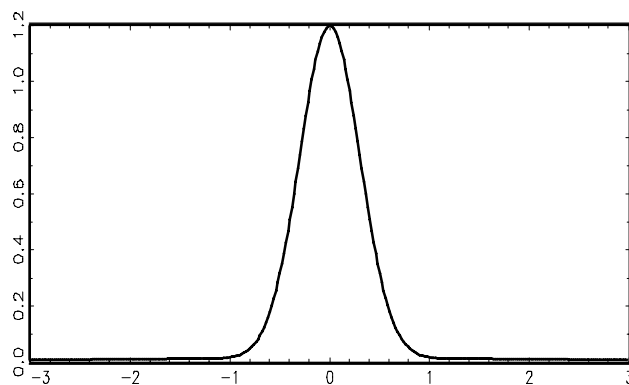
Appendices to

Empirical Bayes Forecasts of One Time Series Using Many Predictors

Thomas Knox
James H. Stock
Mark W. Watson

September, 2000

1 Appendix A: Proofs of Theorems

The first part of this appendix contains proofs of Theorems 1 and 2. The second part contains proofs of Theorems 3 and 4. Please note that when an equation number is referred to in this appendix, the reference is to the equation in this appendix (rather than in the body of the paper) which has that number, unless otherwise stated.

1.1 A.1 Proofs of Theorems 1 and 2

Before we begin, note that Assumption 2 clearly implies the following summability inequalities (due to the exponentially decaying upper bound on the ν_n)

$$
\begin{aligned}
\sum_{m=1}^{\infty} \sum_{n=m}^{\infty} \nu_n &\leq D < \infty \\
\sum_{n=1}^{\infty} n\nu_n &\leq D < \infty
\end{aligned}
\tag{1}
$$

which evidently yield (since ν_n is nonnegative by definition)

$$
\sum_{n=1}^{\infty} \nu_n \leq D < \infty.
\tag{2}
$$

Proof of Theorem 1: First, we show unbiasedness:

$$
\begin{aligned}
E\left[\hat{\sigma}_\varepsilon^2\right] &= \frac{1}{T-K} E\left[\varepsilon'\varepsilon\right] - \frac{1}{T(T-K)} E\left[\varepsilon'XX'\varepsilon\right] \\
&= \frac{T}{T-K} \sigma_\varepsilon^2 - \frac{1}{T(T-K)} \sum_{i=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[X_{is}X_{it}\varepsilon_s\varepsilon_t\right] \\
&= \frac{T}{T-K} \sigma_\varepsilon^2 - \frac{\sigma_\varepsilon^2}{T(T-K)} \sum_{i=1}^{K} E\left[\sum_{t=1}^{T} X_{it}^2\right]
\end{aligned}
\tag{3}
$$

A-1

$$= \frac{T}{T-K}\sigma_\varepsilon^2 - \frac{KT\sigma_\varepsilon^2}{T(T-K)}$$

$$= \sigma_\varepsilon^2$$

where the first equality follows easily from the definition of $\hat{\sigma}_\varepsilon^2$, the second is by Assumption 1 and calculation, for the first and second terms respectively, the third equality is due to the fact that $E[X_{is}X_{it}\varepsilon_s\varepsilon_t] = 0$ if $s \neq t$, and $= \sigma_\varepsilon^2 E[X_{it}^2]$ if $s = t$. The fourth equality follows from the fact that $\sum_{t=1}^T X_{it}^2 = T$ and from trivial summation. The final equality follows by simple cancellation. Note that the unbiasedness does *not* depend on the mixing assumption, although it is crucial later in the proof.

Before entering into the proof of the squared-error bound, it is worth noting that, for any random variables $Z_1, \ldots, Z_m$ each of which has an nth absolute moment, and if $p_i \geq 0$, $\sum_{i=1}^m p_i = n$, then, by iterating Hölder's inequality,

$$|E[\Pi_{i=1}^m Z_i^{p_i}]| \leq \Pi_{i=1}^m (E[|Z_i|^n])^{p_i / n}. \tag{4}$$

This fact will prove useful when we apply Assumption 1 to various cross-moments below, and we will not explicitly cite it when it is applied, in the interest of brevity.

Now, using unbiasedness,

$$E\left[\left(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\right)^2\right] \tag{5}$$

$$= E\left[\left(\hat{\sigma}_\varepsilon^2\right)^2\right] - \sigma_\varepsilon^4$$

$$= E\left[\frac{1}{(T-K)^2}\left(\varepsilon'\left(I-\frac{XX'}{T}\right)\varepsilon\right)\left(\varepsilon'\left(I-\frac{XX'}{T}\right)\varepsilon\right)\right] - \sigma_\varepsilon^4$$

$$= E\left[\frac{1}{(T-K)^2}\left[\begin{array}{c}(\varepsilon'\varepsilon)^2 - 2\left(\varepsilon'\varepsilon\right)\left(\varepsilon'\frac{XX'}{T}\varepsilon\right) + \\ \left(\varepsilon'\frac{XX'}{T}\varepsilon\right)\left(\varepsilon'\frac{XX'}{T}\varepsilon\right)\end{array}\right]\right] - \sigma_\varepsilon^4$$

$$= \left\{\begin{array}{c}E\left[\frac{1}{(T-K)^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\varepsilon_s^2\varepsilon_t^2\right] \\[2mm] +E\left[\frac{-2}{T(T-K)^2}\sum_{t=1}^{T}\sum_{i=1}^{K}\sum_{s=1}^{T}\sum_{u=1}^{T}\varepsilon_t^2 X_{is}X_{iu}\varepsilon_s\varepsilon_u\right] \\[2mm] +E\left[\begin{array}{c}\frac{1}{T^2(T-K)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{u=1}^{T}\sum_{w=1}^{T} \\ X_{is}X_{it}X_{ju}X_{jw}\varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_w\end{array}\right]\end{array}\right\}$$
$$-\sigma_\varepsilon^4$$

and we shall address each of the three expectations in the final expression in turn.

The first expectation can be written as

$$E\left[\frac{1}{(T-K)^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\varepsilon_s^2\varepsilon_t^2\right] = \frac{1}{(T-K)^2}\sum_{s=1}^{T}\sum_{t=1}^{T}E\left[\varepsilon_s^2\varepsilon_t^2\right] \tag{6}$$

$$= \frac{1}{(T-K)^2}\sum_{s=1}^{T}\left[(T-1)\sigma_\varepsilon^4 + E\left[\varepsilon_s^4\right]\right]$$

$$= \frac{T(T-1)}{(T-K)^2}\sigma_\varepsilon^4 + \frac{\sum_{s=1}^{T}E\left[\varepsilon_s^4\right]}{(T-K)^2}.$$

so that

$$\left|E\left[\frac{1}{(T-K)^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\varepsilon_s^2\varepsilon_t^2\right] - \frac{T(T-1)}{(T-K)^2}\sigma_\varepsilon^4\right| \tag{7}$$

$$\leq \frac{\sum_{s=1}^{T}E\left[\varepsilon_s^4\right]}{(T-K)^2}$$

$$\leq \frac{DT}{(T-K)^2}$$

$$\leq \frac{C_1}{K}$$

A-3

where the first inequality is clear, the second follows from Assumption 1, and the third is by the asymptotic nesting which has been assumed, in which K is an asymptotically constant fraction of T.

The second expectation may be evaluated as follows:

$$E\left[\frac{-2}{T\left(T-K\right)^2}\sum_{t=1}^{T}\sum_{i=1}^{K}\sum_{s=1}^{T}\sum_{u=1}^{T}\varepsilon_t^2 X_{is}X_{iu}\varepsilon_s\varepsilon_u\right] \tag{8}$$

$$=\frac{-2}{T\left(T-K\right)^2}\sum_{t=1}^{T}\sum_{i=1}^{K}\sum_{s=1}^{T}\sum_{u=1}^{T}E\left[\varepsilon_t^2 X_{is}X_{iu}\varepsilon_s\varepsilon_u\right]$$

Now we must break out 6 cases:

Case 1: $s = t = u$.

This contributes, in *absolute value*,

$$\frac{2}{T\left(T-K\right)^2}\sum_{i=1}^{K}\sum_{t=1}^{T}E\left[X_{it}^2\varepsilon_t^4\right] \tag{9}$$

$$\leq \frac{2D}{T\left(T-K\right)^2}\sum_{i=1}^{K}\sum_{t=1}^{T}E\left[X_{it}^2\right]$$

$$= \frac{2D}{T\left(T-K\right)^2}\sum_{i=1}^{K}E\left[\sum_{t=1}^{T}X_{it}^2\right]$$

$$= \frac{2KTD}{T\left(T-K\right)^2}$$

$$\leq \frac{C_2}{K}$$

where the first inequality is by Assumption 1, the first equality is by the linearity of expectations, the second equality is by $\sum_{t=1}^{T}X_{it}^2 = T$, and the second inequality is by the asymptotic nesting, in which K is asymptotically a constant fraction of T.

A-4

Case 2: All time subscripts are distinct.

Then $E\left[X_{is}X_{iu}\varepsilon_s\varepsilon_u\varepsilon_t^2\right] = 0$ by using the m. d. s. property of ε, or, if t is the greatest subscript, by the homoskedastic m. d. s. property of ε followed by the m. d. s. property.

Case 3: $s = t > u$ or $u = t > s$.

Suppose w. l. o. g. that $s = t > u$; then

$$\left| E\left[X_{it}\varepsilon_t^3 X_{iu}\varepsilon_u\right]\right| \tag{10}$$

$$= \left| E\left[X_{it}\varepsilon_t^3 X_{iu}\varepsilon_u\right] - E\left[X_{it}\varepsilon_t^3\right] E\left[X_{iu}\varepsilon_u\right]\right|$$

$$\leq \nu_{t-u}\left(E\left[X_{it}^2\varepsilon_t^6\right]\right)^{1/2}\left(E\left[X_{iu}^2\varepsilon_u^2\right]\right)^{1/2}$$

$$\leq M_1\nu_{t-u} \tag{11}$$

where the equality follows from the fact that $E\left[X_{iu}\varepsilon_u\right] = 0$, and the first inequality follows from Doukhan (1994, Theorem 3 (5) on page 9), which states that, if r and z are $\mathcal{H}_1^m$-measurable and $\mathcal{H}_{m+n}^\infty$-measurable random variables, respectively, and if $E\left[r^2\right]$, $E\left[z^2\right] < \infty$, then

$$\left|Cov\left(r,z\right)\right| \leq \nu_n\sqrt{E\left[r^2\right]}\sqrt{E\left[z^2\right]}.$$

Note in particular that the moment bounds hold by Assumption 1, and that the σ-field $\mathcal{H}_a^b$ is generated by the random variables $\{X_a, \varepsilon_a, \ldots, X_b, \varepsilon_b\}$. The second

inequality follows from the uniform (over i, u, and t) bounds on the moments guaranteed by Assumption 1. Thus, the absolute value of the contribution of these terms to the expectation is

$$\frac{2}{T(T-K)^2}\left|\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{t=u+1}^{T}E\left[X_{it}\varepsilon_t^3 X_{iu}\varepsilon_u\right]\right| \tag{12}$$

$$\leq \frac{2}{T(T-K)^2}\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{t=u+1}^{T}\left|E\left[X_{it}\varepsilon_t^3 X_{iu}\varepsilon_u\right]\right|$$

$$\leq \frac{2M_1}{T(T-K)^2}\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{t=u+1}^{T}\nu_{t-u}$$

$$= \frac{2M_1}{T(T-K)^2}\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{n=1}^{T-u}\nu_n$$

$$\leq \frac{2M_1}{T(T-K)^2}\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{n=1}^{\infty}\nu_n$$

$$\leq \frac{2M_1 C^*}{T(T-K)^2}\sum_{i=1}^{K}\sum_{u=1}^{T}1$$

$$= \frac{2M_1 C^* KT}{T(T-K)^2}$$

$$\leq \frac{C_3}{K}$$

where the first inequality is by the triangle inequality, the second inequality is by the preceding display, the first equality is by setting $n = t - u$, the third inequality is due to the fact that $\nu_n \geq 0$ by definition, the fourth inequality is by expression (2), and second equality is by trivial summation, and the final inequality is due to the asymptotic nesting, in which K is asymptotically a constant fraction of T.

Case 4: $s = t < u$.

A-6

Here $E\left[X_{it}\varepsilon_t^3 X_{iu}\varepsilon_u\right] = 0$ by the m. d. s. property of ε.

Case 5: $s = u > t$.

Here $E\left[X_{is}^2\varepsilon_s^2\varepsilon_t^2\right] = \sigma_\varepsilon^2 E\left[X_{is}^2\varepsilon_t^2\right]$ by the homoskedastic m. d. s. property of ε, and

$$\left|E\left[X_{is}^2\varepsilon_t^2\right] - E\left[X_{is}^2\right]E\left[\varepsilon_t^2\right]\right| \tag{13}$$

$$\leq \nu_{s-t}\left(E\left[X_{is}^4\right]\right)^{1/2}\left(E\left[\varepsilon_t^4\right]\right)^{1/2}$$

$$\leq M_2\nu_{s-t}$$

where the first inequality is that of Doukhan (1994, Theorem 3 (5) on page 9) used above (noting that Assumption 1 guarantees moment existence), while the second follows from Assumption 1's uniform (over i, s, and t) bound on the moments. But $E\left[X_{is}^2\right]E\left[\varepsilon_t^2\right] = \sigma_\varepsilon^2 E\left[X_{is}^2\right]$. Thus

$$\left|\begin{array}{l}\frac{-2}{T(T-K)^2}\sum_{i=1}^K\sum_{t=1}^T\sum_{s=t+1}^T E\left[X_{is}^2\varepsilon_s^2\varepsilon_t^2\right]\\[2mm] -\left(\frac{-2\sigma_\varepsilon^4}{T(T-K)^2}\sum_{i=1}^K\sum_{t=1}^T\sum_{s=t+1}^T E\left[X_{is}^2\right]\right)\end{array}\right| \tag{14}$$

$$\leq \frac{2}{T(T-K)^2}\sum_{i=1}^K\sum_{t=1}^T\sum_{s=t+1}^T\left|E\left[X_{is}^2\varepsilon_s^2\varepsilon_t^2\right] - \sigma_\varepsilon^4 E\left[X_{is}^2\right]\right|$$

$$\leq \frac{2M_2}{T(T-K)^2}\sum_{i=1}^K\sum_{t=1}^T\sum_{s=t+1}^T\nu_{s-t}$$

$$\leq \frac{2M_2}{T(T-K)^2}\sum_{i=1}^K\sum_{t=1}^T\sum_{n=1}^\infty\nu_n$$

$$\leq \frac{2M_2C^*}{T(T-K)^2}\sum_{i=1}^K\sum_{t=1}^T 1$$

A-7

$$= \frac{2M_2 C^* K T}{T (T - K)^2}$$

$$\leq \frac{C_4}{K}.$$

Case 6: $t > s = u$.

Here $E\left[X_{is}^2 \varepsilon_s^2 \varepsilon_t^2\right] = \sigma_\varepsilon^4 E\left[X_{is}^2\right]$ by the homoskedastic m. d. s. property of ε as

given in Assumption 1. Thus the overall contribution of these terms is:

$$\frac{-2\sigma_\varepsilon^4}{T(T-K)} \sum_{i=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{t-1} E\left[X_{is}^2\right]. \tag{15}$$

Now that we have broken out the cases, we may pull them back together again:

$$\left| E\left[\frac{-2}{T(T-K)^2} \sum_{t=1}^{T} \sum_{i=1}^{K} \sum_{s=1}^{T} \sum_{u=1}^{T} \varepsilon_t^2 X_{is} X_{iu} \varepsilon_s \varepsilon_u\right] - \left(\frac{-2KT}{(T-K)^2}\sigma_\varepsilon^2\right) \right| \tag{16}$$

$$\leq \frac{C_1 + C_2 + C_3 + C_4}{K} + \left| \begin{array}{c} \frac{-2\sigma_\varepsilon^4}{T(T-K)^2} \sum_{i=1}^{K} \sum_{t=1}^{T} \sum_{s=t+1}^{T} E\left[X_{is}^2\right] \\ + \frac{-2\sigma_\varepsilon^4}{T(T-K)} \sum_{i=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{t-1} E\left[X_{is}^2\right] \\ - \left(\frac{-2KT}{(T-K)^2}\sigma_\varepsilon^2\right) \end{array} \right|$$

$$\leq \frac{C_1 + C_2 + C_3 + C_4}{K} + \frac{2\sigma_\varepsilon^4}{T(T-K)^2} \sum_{i=1}^{K} \sum_{t=1}^{T} E\left[X_{it}^2\right] +$$

$$\left| \frac{-2\sigma_\varepsilon^4}{T(T-K)^2} \sum_{i=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[X_{is}^2\right] - \left(\frac{-2KT}{(T-K)^2}\sigma_\varepsilon^2\right) \right|$$

$$= \frac{C_1 + C_2 + C_3 + C_4}{K} + \frac{2\sigma_\varepsilon^4}{T(T-K)^2} \sum_{i=1}^{K} E\left[\sum_{t=1}^{T} X_{it}^2\right] +$$

$$\left| \frac{-2\sigma_\varepsilon^4}{T(T-K)^2} \sum_{i=1}^{K} \sum_{t=1}^{T} E\left[\sum_{s=1}^{T} X_{is}^2\right] - \left(\frac{-2KT}{(T-K)^2}\sigma_\varepsilon^2\right) \right|$$

$$= \frac{C_1 + C_2 + C_3 + C_4}{K} + \frac{2\sigma_\varepsilon^4 K}{(T-K)^2} +$$

A-8

$$
\begin{aligned}
& \left| \frac{-2\sigma_\varepsilon^4 KT^2}{T(T-K)^2} - \left(\frac{-2KT}{(T-K)^2} \sigma_\varepsilon^2 \right) \right| \\
& = \quad \frac{C_1 + C_2 + C_3 + C_4}{K} + \frac{2\sigma_\varepsilon^4 K}{(T-K)^2} \\
& \leq \quad \frac{C_5}{K}
\end{aligned}
$$

Finally, the third expectation is

$$
\begin{aligned}
E & \left[\begin{array}{c} \frac{1}{T^2(T-K)^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{s=1}^T \sum_{t=1}^T \sum_{u=1}^T \sum_{w=1}^T \\[2mm] X_{is} X_{it} X_{ju} X_{jw} \varepsilon_s \varepsilon_t \varepsilon_u \varepsilon_w \end{array} \right] \qquad (17) \\
& = \quad \frac{1}{T^2(T-K)^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{s=1}^T \sum_{t=1}^T \sum_{u=1}^T \sum_{w=1}^T \\
& \quad E\left[X_{is} X_{it} X_{ju} X_{jw} \varepsilon_s \varepsilon_t \varepsilon_u \varepsilon_w \right]
\end{aligned}
$$

for which we will need to consider 11 different cases.

Case 1: $s = t = u = w$.

$E\left[X_{is}^2 X_{js}^2 \varepsilon_s^4 \right] \leq DE\left[X_{is}^2 X_{js}^2 \right] \leq D^2$ by Assumption 1 (and nonnegativity, for

the first inequality), so the absolute value of the contribution of these terms is

$$
\begin{aligned}
& \frac{1}{T^2(T-K)^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{s=1}^T E\left[X_{is}^2 X_{js}^2 \varepsilon_s^4 \right] \qquad (18) \\
& \leq \quad \frac{D^2}{T^2(T-K)^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{s=1}^T 1 \\
& = \quad \frac{K^2 D^2}{T(T-K)^2} \\
& \leq \quad \frac{C_6}{K}.
\end{aligned}
$$

Case 2: All time subscripts are distinct.

Here $E\left[X_{is}X_{it}X_{ju}X_{jw}\varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_w\right] = 0$ by the m. d. s. property of ε (Assumption 1).

Case 3: $s = t > u > w$ (w. l. o. g.; also, $t = u > w > s$, etc.).

The terms here are of the form $E\left[X_{it}^2\varepsilon_t^2 X_{ju}\varepsilon_u X_{jw}\varepsilon_w\right]$. Noting that, by the m. d. s. property of ε (Assumption 1), $E\left[X_{ju}\varepsilon_u X_{jw}\varepsilon_w\right] = 0$, we see that

$$\left|E\left[X_{it}^2\varepsilon_t^2 X_{ju}\varepsilon_u X_{jw}\varepsilon_w\right]\right| \tag{19}$$

$$\leq \left|E\left[X_{it}^2\varepsilon_t^2 X_{ju}\varepsilon_u X_{jw}\varepsilon_w\right] - E\left[X_{it}^2\varepsilon_t^2\right]E\left[X_{ju}\varepsilon_u X_{jw}\varepsilon_w\right]\right|$$

$$\leq \nu_{t-u}\left(E\left[X_{it}^4\varepsilon_t^4\right]\right)^{1/2}\left(E\left[X_{ju}^2\varepsilon_u^2 X_{jw}^2\varepsilon_w^2\right]\right)^{1/2}$$

$$\leq M_3\nu_{t-u}$$

where the first inequality follows by the observation preceding the display, the second inequality is by Doukhan (1994, Theorem 3 (5) on page 9) and Assumption 2, and the third inequality follows from Assumption 1.

Observing that $E\left[X_{jw}\varepsilon_w\right] = 0$ by the m. d. s. property of ε (Assumption 1), we also obtain

$$\left|E\left[X_{it}^2\varepsilon_t^2 X_{ju}\varepsilon_u X_{jw}\varepsilon_w\right]\right| \tag{20}$$

$$\leq \left|E\left[X_{it}^2\varepsilon_t^2 X_{ju}\varepsilon_u X_{jw}\varepsilon_w\right] - E\left[X_{it}^2\varepsilon_t^2 X_{ju}\varepsilon_u\right]E\left[X_{jw}\varepsilon_w\right]\right|$$

$$\leq \quad \nu_{u-w} \left(E\left[X_{it}^4 \varepsilon_t^4 X_{ju}^2 \varepsilon_u^2 \right] \right)^{1/2} \left(E\left[X_{jw}^2 \varepsilon_w^2 \right] \right)^{1/2}$$

$$\leq \quad M_4 \nu_{u-w}$$

in an entirely similar fashion.

Thus, the total absolute-value of the contribution of the terms covered by this case satisfies

$$\left| \frac{1}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{w=1}^{T} \sum_{u=w+1}^{T} \sum_{t=u+1}^{T} E\left[X_{it}^2 \varepsilon_t^2 X_{ju} \varepsilon_u X_{jw} \varepsilon_w \right] \right| \tag{21}$$

$$\leq \quad \frac{1}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{w=1}^{T} \sum_{u=w+1}^{T} \sum_{t=u+1}^{T} \left| E\left[X_{it}^2 \varepsilon_t^2 X_{ju} \varepsilon_u X_{jw} \varepsilon_w \right] \right|$$

$$\leq \quad \frac{1}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{w=1}^{T} \sum_{u=w+1}^{T} \sum_{t=u+1}^{T} \min \left\{ M_3 \nu_{t-u}, M_4 \nu_{u-w} \right\}$$

$$\leq \quad \frac{\max \{ M_3, M_4 \}}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{w=1}^{T} \sum_{u=w+1}^{T} \sum_{t=u+1}^{T} \min \left\{ \nu_{t-u}, \nu_{u-w} \right\}$$

$$\leq \quad \frac{D \max \{ M_3, M_4 \}}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{w=1}^{T} \sum_{u=w+1}^{T} \sum_{t=u+1}^{T} \min \left\{ e^{-\frac{\lambda}{2}(t-u)}, e^{-\frac{\lambda}{2}(u-w)} \right\}$$

$$= \quad \frac{D \max \{ M_3, M_4 \}}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{w=1}^{T} \sum_{u=w+1}^{T} \left\{ \sum_{t=u+1}^{2u-w} e^{-\frac{\lambda}{2}(u-w)} + \sum_{t=2u-w+1}^{T} e^{-\frac{\lambda}{2}(t-u)} \right\}$$

$$= \quad \frac{D \max \{ M_3, M_4 \}}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{w=1}^{T} \left\{ \begin{array}{l} \sum_{u=w+1}^{T} (u-w) e^{-\frac{\lambda}{2}(u-w)} \\[2mm] + \sum_{u=w+1}^{T} \sum_{t=2u-w+1}^{T} e^{-\frac{\lambda}{2}(t-u)} \end{array} \right\}$$

$$= \quad \frac{D \max \{ M_3, M_4 \}}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{w=1}^{T} \left\{ \begin{array}{l} \sum_{n=1}^{T-w} n e^{-\frac{\lambda}{2} n} \\[2mm] + \sum_{n=1}^{T-w} \sum_{m=n+1}^{T-(n+w)} e^{-\frac{\lambda}{2} m} \end{array} \right\}$$

$$\leq \quad \frac{D \max \{ M_3, M_4 \}}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{w=1}^{T} \left\{ \begin{array}{l} \sum_{n=1}^{\infty} n e^{-\frac{\lambda}{2} n} \\[2mm] + \sum_{n=1}^{\infty} \sum_{m=n+1}^{\infty} e^{-\frac{\lambda}{2} m} \end{array} \right\}$$

A-11

$$\leq \frac{2D^2 \max\{M_3, M_4\}}{T^2 (T - K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{w=1}^{T} 1$$

$$= \frac{2D^2 \max\{M_3, M_4\} K^2}{T (T - K)^2}$$

$$\leq \frac{C_7}{K}$$

where the first inequality is clear, the second inequality follows from the two bounds given above (since each of the previous two bounds holds for each term, the smaller of them must hold), the third inequality is evident, the fourth inequality follows from Assumption 2, the first equality follows from the fact that $t - u > u - w$ if and only if $t > 2u - w$ (and the monotonicity of e^{-n}), the second equality follows from trivial summation and rearrangement, the third equality follows upon setting $m = t - u$ and $n = u - w$, the fifth inequality follows from the nonnegativity of ν_n and n, the sixth inequality follows directly from the bounds of expression (1), the third equality is due to simple summation, and the seventh and final inequality follows from the asymptotic nesting, in which K is asymptotically a constant fraction of T.

Case 4: $s > t = u > w$ (w. l. o. g.; this case includes all terms whose time subscripts take this form, with one greatest, two equal, and one least).

Here we obtain, by the m. d. s. property of ε (Assumption 1), $E\left[X_{is}\varepsilon_s X_{it} X_{jt} \varepsilon_t^2 X_{jw} \varepsilon_w\right] = 0$.

Case 5: $s > t > u = w$ (w. l. o. g.; this case includes all terms whose time subscripts take this form, with one greatest, one intermediate, and two equal and minimal).

Just as in Case 4, $E\left[X_{is}\varepsilon_s X_{it}\varepsilon_t X_{ju}^2 \varepsilon_u^2\right] = 0$ by the m. d. s. property of ε.

Case 6: $s = t = u > w$ (w. l. o. g.; this case includes all terms which have three equal time subscripts and one lesser time subscript).

Noting that $E\left[X_{jw}\varepsilon_w\right] = 0$ by the m. d. s. property of ε (Assumption 1), we have that

$$\left| E\left[X_{it}^2 X_{jt}\varepsilon_t^3 X_{jw}\varepsilon_w\right] \right| \tag{22}$$

$$\leq \left| E\left[X_{it}^2 X_{jt}\varepsilon_t^3 X_{jw}\varepsilon_w\right] - E\left[X_{it}^2 X_{jt}\varepsilon_t^3\right] E\left[X_{jw}\varepsilon_w\right] \right|$$

$$\leq \nu_{t-w}\left(E\left[X_{it}^4 X_{jt}^2 \varepsilon_t^6\right]\right)^{1/2}\left(E\left[X_{jw}^2 \varepsilon_w^2\right]\right)^{1/2}$$

$$\leq M_5 \nu_{t-w}$$

where the first inequality follows from the preceding observation, the second is due to Doukhan (1994, Theorem 3 (5) on page 9) and Assumption 2, and the final inequality is due to the uniform moment bounds of Assumption 1.

The total absolute-value contribution of the terms handled in this case is thus

$$\left| \frac{1}{T^2 (T-K)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T}\sum_{w=t+1}^{T} E\left[X_{it}^2 X_{jt}\varepsilon_t^3 X_{jw}\varepsilon_w\right] \right| \tag{23}$$

A-13

$$\leq \frac{1}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{w=t+1}^{T} \left| E\left[X_{it}^2 X_{jt} \varepsilon_t^3 X_{jw} \varepsilon_w \right] \right|$$

$$\leq \frac{M_5}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{w=t+1}^{T} \nu_{t-w}$$

$$\leq \frac{DM_5}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} 1$$

$$= \frac{DM_5 K^2}{T (T-K)^2}$$

$$\leq \frac{C_8}{K}$$

where the first inequality is standard, the second follows from the above bound on each summand, the third inequality follows from the summability condition of expression (2), the equality follows by simple summation, and the final inequality is due to the asymptotic nesting, in which K is a constant fraction of T asymptotically.

Case 7: $s > t = u = w$ (w. l. o. g.; this case handles all terms which have three equal time subscripts and one greater time subscript).

By the m. d. s. property of the ε, we have that $E\left[X_{is} \varepsilon_s X_{it} X_{jt}^2 \varepsilon_t^3 \right] = 0$.

Case 8: $s = t > u = w$ (note that here, we deal only with the specific subscript ordering given).

We have that $E\left[X_{it}^2 \varepsilon_t^2 X_{ju}^2 \varepsilon_u^2 \right] = \sigma_\varepsilon^2 E\left[X_{it}^2 X_{ju}^2 \varepsilon_u^2 \right]$, $E\left[X_{it}^2 \varepsilon_t^2 \right] = \sigma_\varepsilon^2 E\left[X_{it}^2 \right]$, and $E\left[X_{ju}^2 \varepsilon_u^2 \right] = \sigma_\varepsilon^2 E\left[X_{ju}^2 \right]$, so

$$\left| E\left[X_{it}^2 \varepsilon_t^2 X_{ju}^2 \varepsilon_u^2 \right] - \sigma_\varepsilon^4 E\left[X_{it}^2 \right] E\left[X_{ju}^2 \right] \right| \tag{24}$$

A-14

$$= \sigma_\varepsilon^2 \left| E\left[X_{it}^2 X_{ju}^2 \varepsilon_u^2\right] - \sigma_\varepsilon^2 E\left[X_{it}^2\right] E\left[X_{ju}^2\right] \right|$$

$$= \sigma_\varepsilon^2 \left| E\left[X_{it}^2 X_{ju}^2 \varepsilon_u^2\right] - E\left[X_{it}^2\right] E\left[X_{ju}^2 \varepsilon_u^2\right] \right|$$

$$\leq \sigma_\varepsilon^2 \nu_{t-u} \left(E\left[X_{it}^4\right]\right)^{1/2} \left(E\left[X_{ju}^4 \varepsilon_u^4\right]\right)^{1/2}$$

$$\leq M_6 \nu_{t-u}$$

where the first two equalities follow from the observations made immediately above the display, the first inequality comes from Hall and Heyde (1980, Theorem A.6 on page 278) and Assumption 2, and the final inequality comes from the uniform moment bounds given by Assumption 1.

We can now bound the absolute value of the difference between the total contribution of the terms handled by this case and

$$\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{u=1}^T \sum_{t=u+1}^T E\left[X_{it}^2\right] E\left[X_{ju}^2\right]. \tag{25}$$

This is done as follows:

$$\left| \begin{array}{c} \frac{1}{T^2(T-K)^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{u=1}^T \sum_{t=u+1}^T E\left[X_{it}^2 \varepsilon_t^2 X_{ju}^2 \varepsilon_u^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{u=1}^T \sum_{t=u+1}^T E\left[X_{it}^2\right] E\left[X_{ju}^2\right] \end{array} \right| \tag{26}$$

$$\leq \frac{1}{T^2(T-K)^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{u=1}^T \sum_{t=u+1}^T \left| E\left[X_{it}^2 \varepsilon_t^2 X_{ju}^2 \varepsilon_u^2\right] - \sigma_\varepsilon^4 E\left[X_{it}^2\right] E\left[X_{ju}^2\right] \right|$$

$$\leq \frac{M_6}{T^2(T-K)^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{u=1}^T \sum_{t=u+1}^T \nu_{t-u}$$

$$\leq \frac{M_6}{T^2(T-K)^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{u=1}^T \sum_{n=1}^\infty \nu_n$$

$$\leq \frac{M_6 D}{T^2(T-K)^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{u=1}^T 1$$

$$
\begin{aligned}
&= \frac{M_6 D K^2}{T\left(T-K\right)^2} \\
&\leq \frac{C_9}{K}
\end{aligned}
$$

where the first inequality is standard, the second is by the above bound on each summand, the third follows from nonnegativity of the ν_n and setting $n = t - u$, the fourth follows from expression (2), the equality is by simple summation, and the final inequality is due to the asymptotic nesting, in which K is an asymptotically constant fraction of T.

Case 9: $u = w > s = t$ (note that here, we deal only with the specific subscript ordering given).

This is identical to Case 8 above, except that we get the time terms $t < u$ rather than the terms $t > u$. Thus, the result is that the total contribution of the terms to which this case applies is close to

$$
\frac{\sigma_\varepsilon^4}{T^2\left(T-K\right)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{u=1}^{T}\sum_{t=1}^{u-1} E\left[X_{it}^2\right] E\left[X_{ju}^2\right] \tag{27}
$$

in the sense of the following bound:

$$
\left|
\begin{array}{l}
\frac{1}{T^2(T-K)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{u=1}^{T}\sum_{t=1}^{u-1} E\left[X_{it}^2\varepsilon_t^2 X_{ju}^2\varepsilon_u^2\right] \\[2mm]
-\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{u=1}^{T}\sum_{t=1}^{u-1} E\left[X_{it}^2\right] E\left[X_{ju}^2\right]
\end{array}
\right| \tag{28}
$$
$$
\leq \frac{C_{10}}{K}.
$$

A-16

Case 10: $s = u > t = w$ (note that here, we deal only with the specific ordering of the subscripts given).

$E\left[X_{is}X_{js}\varepsilon_s^2 X_{it}X_{jt}\varepsilon_t^2\right] = \sigma_\varepsilon^2 E\left[X_{is}X_{js}X_{it}X_{jt}\varepsilon_t^2\right]$ by the homoskedastic m. d. s. property of ε (from Assumption 1). Now, $E\left[X_{it}X_{jt}\varepsilon_t^2\right] = \sigma_\varepsilon^2 E\left[X_{it}X_{jt}\right]$ for the same reason, so

$$
\left| E\left[X_{is}X_{js}\varepsilon_s^2 X_{it}X_{jt}\varepsilon_t^2\right] - \sigma_\varepsilon^4 E\left[X_{it}X_{jt}\right]E\left[X_{is}X_{js}\right]\right| \tag{29}
$$

$$
= \left| \sigma_\varepsilon^2 E\left[X_{is}X_{js}X_{it}X_{jt}\varepsilon_t^2\right] - \sigma_\varepsilon^2 E\left[X_{it}X_{jt}\varepsilon_t^2\right]E\left[X_{is}X_{js}\right]\right|
$$

$$
= \sigma_\varepsilon^2 \left| E\left[X_{is}X_{js}X_{it}X_{jt}\varepsilon_t^2\right] - E\left[X_{it}X_{jt}\varepsilon_t^2\right]E\left[X_{is}X_{js}\right]\right|
$$

$$
\leq \sigma_\varepsilon^2 \nu_{s-t}\left(E\left[X_{it}^2 X_{jt}^2 \varepsilon_t^4\right]\right)^{1/2}\left(E\left[X_{is}^2 X_{js}^2\right]\right)^{1/2}
$$

$$
\leq M_7 \nu_{s-t}
$$

where the first two equalities are by the identities noted immediately above the display, the first inequality is by Doukhan (1994, Theorem 3 (5) on page 9) and Assumption 2, and the final inequality is by the uniform moment bounds given in Assumption 1. Thus, we can obtain, exactly as in Cases 8 and 9 above, a bound on the absolute value of the difference between the total contribution of the terms handled here and the object

$$
\frac{\sigma_\varepsilon^4}{T^2(T-K)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T}\sum_{s=t+1}^{T} E\left[X_{it}X_{jt}\right]E\left[X_{is}X_{js}\right] \tag{30}
$$

A-17

where the bound is:

$$
\left|
\begin{array}{l}
\frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=t+1}^{T} E\left[X_{is}X_{js}\varepsilon_s^2 X_{it}X_{jt}\varepsilon_t^2\right] \\
-\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=t+1}^{T} E\left[X_{it}X_{jt}\right] E\left[X_{is}X_{js}\right]
\end{array}
\right|
$$
$$
\leq \quad \frac{C_{11}}{K}.
$$
(31)

Case 11: $t = w > s = u$ (note that here, we deal only with the specific ordering of the subscripts given).

This is entirely similar to Case 10, except that the bound is derived on the distance between the total contribution of the terms handled here and

$$
\frac{\sigma_\varepsilon^4}{T^2 (T - K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{t-1} E\left[X_{it}X_{jt}\right] E\left[X_{is}X_{js}\right].
$$
(32)

The bound is:

$$
\left|
\begin{array}{l}
\frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{t-1} E\left[X_{is}X_{js}\varepsilon_s^2 X_{it}X_{jt}\varepsilon_t^2\right] \\
-\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{t-1} E\left[X_{it}X_{jt}\right] E\left[X_{is}X_{js}\right]
\end{array}
\right|
$$
$$
\leq \quad \frac{C_{12}}{K}.
$$
(33)

Now we may pull all the cases back together again to get:

$$
\left|
\begin{array}{c}
\frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{u=1}^{T} \sum_{w=1}^{T} \\
E\left[X_{is}X_{it}X_{ju}X_{jw}\varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_w\right] \\
-\frac{\sigma_\varepsilon^4 K^2}{(T-K)^2}
\end{array}
\right|
$$
(34)

A-18

$$\leq \quad \frac{C_7 + C_8 + C_9}{K}$$

$$+ \left| \begin{array}{l} \frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{u=1}^{T} \sum_{t=u+1}^{T} E\left[X_{it}^2 \varepsilon_t^2 X_{ju}^2 \varepsilon_u^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{u=1}^{T} \sum_{t=u+1}^{T} E\left[X_{it}^2\right] E\left[X_{ju}^2\right] \end{array} \right|$$

$$+ \left| \begin{array}{l} \frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{u=1}^{T} \sum_{t=1}^{u-1} E\left[X_{it}^2 \varepsilon_t^2 X_{ju}^2 \varepsilon_u^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{u=1}^{T} \sum_{t=1}^{u-1} E\left[X_{it}^2\right] E\left[X_{ju}^2\right] \end{array} \right|$$

$$+ \left| \begin{array}{l} \frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=t+1}^{T} E\left[X_{is} X_{js} \varepsilon_s^2 X_{it} X_{jt} \varepsilon_t^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=t+1}^{T} E\left[X_{it} X_{jt}\right] E\left[X_{is} X_{js}\right] \end{array} \right|$$

$$+ \left| \begin{array}{l} \frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{t-1} E\left[X_{is} X_{js} \varepsilon_s^2 X_{it} X_{jt} \varepsilon_t^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{t-1} E\left[X_{it} X_{jt}\right] E\left[X_{is} X_{js}\right] \end{array} \right|$$

$$+ \left| \begin{array}{l} \frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[X_{it} X_{jt}\right] E\left[X_{is} X_{js}\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} E\left[X_{it} X_{jt}\right] E\left[X_{it} X_{jt}\right] \end{array} \right|$$

$$+ \left| \begin{array}{c} \frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{u=1}^{T} \sum_{t=1}^{T} E\left[X_{it}^2\right] E\left[X_{ju}^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} E\left[X_{it}^2\right] E\left[X_{jt}^2\right] \\ -\frac{\sigma_\varepsilon^4 K^2}{(T-K)^2} \end{array} \right|$$

$$\leq \quad \frac{C_7 + C_8 + C_9 + C_{10} + C_{11} + C_{12}}{K}$$

$$+ \left| \frac{\sigma_\varepsilon^4}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[X_{it} X_{jt}\right] E\left[X_{is} X_{js}\right] \right|$$

$$+ \left| \frac{\sigma_\varepsilon^4}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} E\left[X_{it} X_{jt}\right] E\left[X_{it} X_{jt}\right] \right|$$

$$+ \left| \frac{\sigma_\varepsilon^4}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} E\left[X_{it}^2\right] E\left[X_{jt}^2\right] \right|$$

$$+ \left| \begin{array}{c} \frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{u=1}^{T} \sum_{t=1}^{T} E\left[X_{it}^2\right] E\left[X_{ju}^2\right] \\ -\frac{\sigma_\varepsilon^4 K^2}{(T-K)^2} \end{array} \right|$$

$$\leq \quad \frac{C_7 + C_8 + C_9 + C_{10} + C_{11} + C_{12} + C_{13} + C_{14}}{K}$$

$$+ \left| \frac{\sigma_\varepsilon^4}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[X_{it}X_{jt}\right] E\left[X_{is}X_{js}\right] \right|$$

$$\leq \quad \frac{C_7 + C_8 + C_9 + C_{10} + C_{11} + C_{12} + C_{13} + C_{14}}{K}$$

$$+ \left| \frac{\sigma_\varepsilon^4}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j\neq i}}^{K} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[X_{it}X_{jt}\right] E\left[X_{is}X_{js}\right] \right|$$

$$+ \left| \frac{\sigma_\varepsilon^4}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[X_{it}^2\right] E\left[X_{is}^2\right] \right|$$

$$\leq \quad \frac{C_7 + C_8 + C_9 + C_{10} + C_{11} + C_{12} + C_{13} + C_{14} + C_{15}}{K}$$

$$= \quad \frac{C_{16}}{K}.$$

We may finally combine the bounds we have obtained for each of the three expectations involved, and notice that $\sigma_\varepsilon^4 = \frac{T^2 - 2KT + K^2}{(T-K)^2}\sigma_\varepsilon^4$, to yield:

$$\left| E\left[\left(\hat{\sigma}_\varepsilon^2\right)^2\right] - \sigma_\varepsilon^4 \right| \tag{35}$$

$$\leq \quad \left| E\left[\frac{1}{(T-K)^2} \sum_{s=1}^{T} \sum_{t=1}^{T} \varepsilon_s^2 \varepsilon_t^2 \right] - \frac{T(T-1)}{(T-K)^2}\sigma_\varepsilon^4 \right|$$

$$+ \left| \frac{T}{(T-K)^2}\sigma_\varepsilon^4 \right|$$

$$+ \left| \begin{array}{c} E\left[\frac{-2}{T(T-K)^2} \sum_{t=1}^{T} \sum_{i=1}^{K} \sum_{s=1}^{T} \sum_{u=1}^{T} \varepsilon_t^2 X_{is} X_{iu} \varepsilon_s \varepsilon_u \right] \\ -\left(\frac{-2KT}{(T-K)^2}\sigma_\varepsilon^2 \right) \end{array} \right|$$

$$+ \left| \begin{array}{c} \frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{u=1}^{T} \sum_{w=1}^{T} \\[2mm] E\left[X_{is}X_{it}X_{ju}X_{jw}\varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_w\right] \\[2mm] - \frac{\sigma_\varepsilon^4 K^2}{(T-K)^2} \end{array} \right|$$

$$\leq \quad \frac{C_1}{K} + \left| \frac{T}{(T-K)^2}\sigma_\varepsilon^4 \right| + \frac{C_5}{K} + \frac{C_{16}}{K}$$

$$\leq \quad \frac{C_1}{K} + \frac{C_{17}}{K} + \frac{C_5}{K} + \frac{C_{16}}{K}$$

$$\leq \quad \frac{C}{K}$$

which completes the proof of Theorem 1. Q.E.D.

Proof of Theorem 2: Because we have w. l. o. g. set $H = I_K$,

$$
\begin{aligned}
R\left(\hat{b}, b\right) &= \frac{1}{T}E\left[\left(\hat{b}-b\right)'\left(\hat{b}-b\right)\right] \qquad\qquad (36) \\[2mm]
&= \frac{1}{T}E\left[\left(\frac{1}{\sqrt{T}}X'\varepsilon\right)'\left(\frac{1}{\sqrt{T}}X'\varepsilon\right)\right] \\[2mm]
&= \frac{1}{T^2}E\left[\sum_{i=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{T}X_{it}X_{is}\varepsilon_t\varepsilon_s\right] \\[2mm]
&= \frac{1}{T^2}\sum_{i=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[X_{it}X_{is}\varepsilon_t\varepsilon_s\right] \\[2mm]
&= \frac{1}{T^2}\sum_{i=1}^{K}\sum_{t=1}^{T}E\left[X_{it}^2\varepsilon_t^2\right] \\[2mm]
&= \frac{\sigma_\varepsilon^2}{T^2}\sum_{i=1}^{K}E\left[\sum_{t=1}^{T}X_{it}^2\right] \\[2mm]
&= \frac{\sigma_\varepsilon^2}{T^2}\sum_{i=1}^{K}T \\[2mm]
&= \frac{K}{T}\sigma_\varepsilon^2
\end{aligned}
$$

A-21

$$\rightarrow \quad \rho\sigma_\varepsilon^2$$

where the first four equalities follow from simple calculation, the fifth equality holds because if $s \neq t$, then $E\left[X_{is}X_{it}\varepsilon_s\varepsilon_t\right] = 0$ by the m. d. s. property of ε (from Assumption 1), the sixth equality is due to the fact that $E\left[X_{it}^2\varepsilon_t^2\right] = \sigma_\varepsilon^2 E\left[X_{it}^2\right]$, the seventh equality follows from $\sum_{t=1}^T X_{it}^2 = T$ (since $X'X = TI_K$), the eighth equality follows from simple summation, and the limit is due to the asymptotic nesting we have specified, in which $\frac{K}{T} \rightarrow \rho$ as $T \rightarrow \infty$. All of the other results are immediate consequences of the above calculation. Q. E. D.

1.2   A.2   Proofs of Theorems 3, 4, and 5

We start by collecting some additional definitions.

Definition 1

$$d_K \equiv \sqrt{\frac{\sigma_\varepsilon^2}{64}\log K}$$

$$\hat{b}_{-i} \equiv \left(\hat{b}_1, \ldots, \hat{b}_{i-1}, \hat{b}_{i+1}, \ldots, \hat{b}_K\right)$$

$$\phi(u) \equiv \text{ the univariate normal density with}$$

$$\text{mean } 0 \text{ and variance } \sigma_\varepsilon^2$$

$$f_{iK}(u_i) \equiv \text{ the marginal likelihood of } \hat{b}_i \text{ given } b_i$$

$$= \int_{u_{-i}} f_K(u)\, du_{-i}$$

A-22

$$\bar{f}_K(x) \equiv \frac{1}{K}\sum_{i=1}^{K} f_{iK}(x)$$

$$f_{ijK}\left(\hat{b}_i - b_i, \hat{b}_j - b_j\right) \equiv \text{the likelihood of } \hat{b}_i \text{ and } \hat{b}_j \text{ given } b_i \text{ and } b_j$$

$$= \int_{u_{-(i,j)}} f_K(u)\, du_{-(i.j)}$$

$$m_{iK}\left(\hat{b}_i\right) \equiv \int_{-\infty}^{\infty} f_{iK}\left(\hat{b}_i - b_i\right) dG(b_i)$$

$$m'_{iK}\left(\hat{b}_i\right) \equiv \frac{d}{d\hat{b}_i}\int_{-\infty}^{\infty} f_{iK}\left(\hat{b}_i - b_i\right) dG(b_i)$$

$$\bar{m}_K(x) \equiv \frac{1}{K}\sum_{i=1}^{K} m_{iK}(x)$$

$$m_\phi\left(\hat{b}_1\right) \equiv \int_{-\infty}^{\infty} \phi\left(\hat{b}_1 - b_1\right) dG(b_1)$$

$$m^C_{ijK}\left(\hat{b}_i \mid \hat{b}_j\right) \equiv \frac{m_{ijK}\left(\hat{b}_i, \hat{b}_j\right)}{m_{iK}\left(\hat{b}_j\right)}$$

$$\equiv \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{ijK}\left(\hat{b}_i - b_i, \hat{b}_j - b_j\right) dG(b_i)\, dG(b_j)}{m_{iK}\left(\hat{b}_j\right)}$$

$$\bar{m}^C_{iK}(x \mid y) \equiv \frac{1}{K-1}\sum_{\substack{j=1,\\ j\neq i}}^{K} m^C_{ijK}(x \mid y)$$

Note that Assumption 7 implies the following limits:

$$s_K q_K \rightarrow \infty \tag{37}$$

$$K s_K^8 \rightarrow \infty$$

$$s_K^2 \log K \rightarrow \infty$$

$$\frac{K^{-5/48} \log^2 K}{s_K^2} \rightarrow 0$$

$$\frac{K^{-5/96} h_K \log K}{s_K^2} \rightarrow 0$$

$$\frac{h_K}{s_K} \rightarrow 0$$

A-23

$$q_K^2 \; / \; K \;\; \rightarrow \;\; 0$$

The following lemmas are used in proving the main results. The first three lemmas collect Berry-Esseen-type results about convergence of certain densities and their derivatives to local limits.

Lemma 1 (Berry-Esseen Results for Densities) *Under Assumptions 1, 2, and 3, we have the following local limit rate results:* $\exists$ *finite* C, K_0 *s.t.* $\forall \, K \geq K_0$,

$$\sup_{i,s} |f_{iK}(s) - \phi(s)| \;\; \leq \;\; CK^{-\frac{1}{4}} \log K \tag{38}$$

$$\sup_{i,s} |f'_{iK}(s) - \phi'(s)| \;\; \leq \;\; CK^{-\frac{1}{8}} \log K \tag{39}$$

Proof of Lemma 1: Let $\eta_{it} = \frac{X_{it}\varepsilon_t}{\sigma_\varepsilon}$. Suppose that the sequence $\eta_{i1}, \eta_{i2}, \ldots$ satisfies Conditions A and C of Appendix B, where the constants in those conditions do not depend on i. Then, by Theorem 2 of Appendix B and a simple change of scale (recalling that σ_ε^2 is bounded away from both zero and infinity),

$$\sup_s |f_{iK}(s) - \phi(s)| \;\; \leq \;\; CK^{-\frac{1}{4}} \log K \tag{40}$$

$$\sup_s |f'_{iK}(s) - \phi'(s)| \;\; \leq \;\; CK^{-\frac{1}{8}} \log K$$

for each i, where the constant C does not depend on i. It follows that these inequalities hold uniformly in $i = 1, \ldots, K$, i. e., that (38) and (39) hold. To prove the theorem, it therefore suffices to prove that $\eta_{i1}, \eta_{i2}, \ldots$ satisfy Conditions A and C with constants that do not depend on i.

We first verify that Condition C is satisfied. By Assumption 1, $E[\eta_{it}] = 0$ and

$$
\begin{aligned}
E\left[\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\eta_{it}\right)^2\right] &= \frac{1}{T\sigma_\varepsilon^2}\sum_{s=1}^{T}\sum_{t=1}^{T}E\left[X_{it}X_{is}\varepsilon_t\varepsilon_s\right] \qquad (41)\\
&= \frac{1}{T\sigma_\varepsilon^2}\sum_{t=1}^{T}E\left[X_{it}^2\varepsilon_t^2\right]\\
&= \frac{1}{T}E\left[\sum_{t=1}^{T}X_{it}^2\right]\\
&= 1
\end{aligned}
$$

where the first equality is by definition, the second is by the m. d. s. property of the ε's (Assumption 1), the third is by the homoskedasticity of the ε's (Assumption 1) and the fourth equality is by the orthonormality of the X's.

Also, by Assumption 1,

$$
\begin{aligned}
\sup_i E\left[\eta_{it}^6\right] &= \sup_i E\left[X_{it}^6\varepsilon_t^6\right] \qquad (42)\\
&\leq \sup_i \left(E\left[X_{it}^{12}\right]E\left[\varepsilon_t^{12}\right]\right)^{1/2}\\
&= \left(\sup_i E\left[X_{it}^{12}\right]E\left[\varepsilon_t^{12}\right]\right)^{1/2}\\
&\leq C
\end{aligned}
$$

so the moment conditions in Condition C hold uniformly in i. The condition on the (time series) mixing coefficients in Condition C and the fact that it hold uniformly in i follow directly from Assumption 2. This verifies that Assumptions 1 and 2 imply Condition C.

A-25

It remains to show that Condition A of Appendix B is implied by Assumption 3 uniformly in i. Let

$$\psi_{it}^{X\varepsilon}(s) \equiv \int_{-\infty}^{\infty} e^{is\eta_{it}} p_{it}^{\eta}\left(\eta_{it} \mid \underline{X}^{(t-1)}, \varepsilon^{(t-1)}\right) d\eta_{it} \tag{43}$$

be the characteristic function of η_{it} (conditional on all past X and ε), and define

$$\psi_t^{\varepsilon}(s) \equiv \int_{-\infty}^{\infty} e^{is\varepsilon_t} p_t^{\varepsilon}\left(\varepsilon_t \mid \underline{X}^{(t-1)}, \varepsilon^{(t-1)}\right) d\varepsilon_t \tag{44}$$

to be the characteristic function of ε_t (conditional on all past X and ε).

We need to show that $\exists \alpha > 0, C_0 < \infty$, and $M_0 < \infty$ such that

$$\forall |s| \geq C_0, \ \sup_{it} \left|\psi_{it}^{X\varepsilon}(s)\right| \leq M_0 |s|^{-\alpha}. \tag{45}$$

Set $C_0 = 1$. Note that Assumption 3 implies that

$$\begin{aligned}
\sup_t |\psi_t^{\varepsilon}(s)| &= \sup_t \left|\int_{-\infty}^{\infty} e^{is\varepsilon_t} p_t^{\varepsilon}\left(\varepsilon_t \mid \underline{X}^{(t-1)}, \varepsilon^{(t-1)}\right) d\varepsilon_t\right| \tag{46} \\
&= \sup_t \left|-\frac{1}{is} \int_{-\infty}^{\infty} e^{is\varepsilon_t} \left[\frac{d}{d\varepsilon_t} p_t^{\varepsilon}\left(\varepsilon_t \mid \underline{X}^{(t-1)}, \varepsilon^{(t-1)}\right)\right] d\varepsilon_t\right| \\
&= \sup_t \left|-\frac{1}{s^2} \int_{-\infty}^{\infty} e^{is\varepsilon_t} \left[\frac{d^2}{d\varepsilon_t^2} p_t^{\varepsilon}\left(\varepsilon_t \mid \underline{X}^{(t-1)}, \varepsilon^{(t-1)}\right)\right] d\varepsilon_t\right| \\
&\leq |s|^{-2} \sup_t \int_{-\infty}^{\infty} \left|e^{is\varepsilon_t}\right| \left|\frac{d^2}{d\varepsilon_t^2} p_t^{\varepsilon}\left(\varepsilon_t \mid \underline{X}^{(t-1)}, \varepsilon^{(t-1)}\right)\right| d\varepsilon_t \\
&= |s|^{-2} \sup_t \int_{-\infty}^{\infty} \left|\frac{d^2}{d\varepsilon_t^2} p_t^{\varepsilon}\left(\varepsilon_t \mid \underline{X}^{(t-1)}, \varepsilon^{(t-1)}\right)\right| d\varepsilon_t \\
&\leq M_7 |s|^{-2}
\end{aligned}$$

where the first equality is by definition, the second is by integration by parts, the third is by another integration by parts, the first inequality is obvious, the fourth equality is

A-26

by $\left| e^{-is\varepsilon_t} \right| = 1$, and the final inequality is by $\sup_t \int_{-\infty}^{\infty} \left| \frac{d^2}{d\varepsilon_t^2} p_t^\varepsilon \left(\varepsilon_t \mid \underline{X}^{(t-1)}, \varepsilon^{(t-1)} \right) \right| d\varepsilon_t <$ ∞ according to Assumption 3.

By Feller (1971, page 527) (or simply a short calculation), we have the first inequality in the following display, and the rest follow from (46) and our assumptions: for $|s| \geq 1$,

$$
\begin{aligned}
\sup_{it} \left| \psi_{it}^{X\varepsilon}(s) \right| \quad &\leq \quad \sup_{it} \left| \int_{-\infty}^{\infty} \psi_t^\varepsilon(sx) \, p_{it}^X(x) \, dx \right| \quad\quad\quad\quad\quad\quad\quad\quad\quad (47) \\
&\leq \quad \sup_{it} \left| \int_1^{\infty} \psi_t^\varepsilon(sx) \, p_{it}^X(x) \, dx \right| + \sup_{it} \left| \int_{1/|s|}^1 \psi_t^\varepsilon(sx) \, p_{it}^X(x) \, dx \right| \\
&\quad + \sup_{it} \left| \int_{-1/|s|}^{1/|s|} \psi_t^\varepsilon(sx) \, p_{it}^X(x) \, dx \right| + \sup_{it} \left| \int_{-1}^{-1/|s|} \psi_t^\varepsilon(sx) \, p_{it}^X(x) \, dx \right| \\
&\quad + \sup_{it} \left| \int_{-\infty}^{-1} \psi_t^\varepsilon(sx) \, p_{it}^X(x) \, dx \right| \\
&\leq \quad \int_1^{\infty} \frac{M_7}{|sx|^2} p_{it}^X(x) \, dx \;+\; \int_{1/|s|}^1 \frac{M_6 M_7}{|sx|^2} dx \;+\; \int_{-1/|s|}^{1/|s|} M_6 dx \\
&\quad + \int_{-1}^{-1/|s|} \frac{M_6 M_7}{|sx|^2} dx \;+\; \int_{-\infty}^{-1} \frac{M_7}{|sx|^2} p_{it}^X(x) \, dx \\
&\leq \quad \frac{2M_7}{|s|^2} \;+\; \frac{2M_6}{|s|} \;+\; \frac{2M_7 M_6}{|s|^2} \;+\; \frac{2M_7 M_6}{|s|}
\end{aligned}
$$

so $\alpha = 1$ and we are finished.

Lemma 2 (Berry-Esseen Results for Joint Densities) *Under Assumptions 1, 2, and 3, we have that $\exists$ finite C, K_0 s.t. $\forall K \geq K_0$,*

$$
\sup_{i,j,s,u} |f_{ijK}(s,u) - \phi(s)\phi(u)| \quad \leq \quad CK^{-\frac{1}{6}} \log K \quad\quad\quad (48)
$$

$$
\sup_{i,j,s,u} \left| \frac{\partial}{\partial s} f_{ijK}(s,u) - \phi'(s)\phi(u) \right| \quad \leq \quad CK^{-\frac{1}{12}} \log K \quad\quad\quad (49)
$$

A-27

Proof of Lemma 2: Let $\eta_{ijt} = \frac{1}{\sigma_\varepsilon} \begin{pmatrix} X_{it}\varepsilon_t \\ X_{jt}\varepsilon_t \end{pmatrix}$. If the sequence of bivariate random variables $\eta_{ij1}, \eta_{ij2}, \ldots$ satisfies Conditions A and C in Appendix B uniformly in (i,j), then Theorem 3 of Appendix B holds uniformly in (i,j) and Lemma 2 follows. The argument that Assumptions 1 and 2 imply Condition C for $\eta_{ij1}, \eta_{ij2}, \ldots$ uniformly in (i,j) parallels the corresponding argument in the proof of Lemma 1 and is omitted.

It remains only to show that Assumption 3 implies Condition A of Appendix B uniformly in (i,j). Let

$$\psi_{it}^{X\varepsilon}(s_1, s_2) \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\left(s_1\eta_{ijt}^{(1)} + s_2\eta_{ijt}^{(2)}\right)} p_{ijt}^\eta \left(\eta_{ijt} \mid \underline{X}^{(t-1)}, \varepsilon^{(t-1)}\right) d\eta_{ijt}^{(1)} d\eta_{ijt}^{(2)} \tag{50}$$

be the characteristic function of η_{ijt}. We need to show that $\exists \alpha > 0, C_0 < \infty$, and $M_0 < \infty$ such that

$$\forall |s| \geq C_0, \ \sup_{it} \left| \psi_{it}^{X\varepsilon}(s) \right| \leq M_0 |s|^{-\alpha}. \tag{51}$$

As in the proof of the corresponding part of Lemma 1, choose $C_0 = 1$.

We cannot use the rather simple method of Lemma 1 above to prove (51), because of the possibility that the vectors (s_1, s_2) and (x_i, x_j) might be orthogonal, causing $\frac{1}{(s_1 x_i + s_2 x_j)^2}$ to be undefined (infinite) even when both $|s|$ and $|x|$ are large.

To avoid this problem, define by A the set of points in the plane such that the angle between the vectors s and x is within γ of either $\frac{\pi}{2}$ or $\frac{3\pi}{2}$ (that is, the vectors are "γ-close" to being orthogonal), and the magnitude of the x vector is between $\frac{1}{|s|}$ and $|s|^{1/4}$. Recall that $(s_1 x_i + s_2 x_j)^2 = |s|^2 |x|^2 \cos^2 \theta_{s,x}$, where $\theta_{s,x}$ is the angle between s

A-28

and x. Also, note that the density $p_{ijt}^X(x_i, x_j)$ is bounded by Assumption 3. Finally, recall that $|\cos\gamma|$ dominates a sawtoothed function of γ (draw a line from each zero of $|\cos\gamma|$ to the nearest maximum of $|\cos\gamma|$ to the left of the zero, and another line to the nearest maximum to the right of the zero, and you will have drawn the sawtoothed function). Thus, $\cos^{-2}\left(\frac{\pi}{2} \pm \gamma\right) \leq \frac{\pi^2}{4}\gamma^{-2}$ for $0 < \gamma \leq \frac{\pi}{2}$, and the same expression holds for $\cos^{-2}\left(\frac{3\pi}{2} \pm \gamma\right)$.

Set $C_0 = \left(\frac{\pi}{2}\right)^{-4/3}$, so we will demonstrate that the inequality in display (51) holds for s such that $|s| \geq \left(\frac{\pi}{2}\right)^{-4/3}$, which implies that $|s|^{-3/4} \leq \frac{\pi}{2}$. Set $\gamma = |s|^{-3/4}$, so that $\gamma \leq \frac{\pi}{2}$ for $|s| \geq C_0$ and we may apply the cosine inequality developed in the preceding paragraph. Let B_r denote a ball of radius r centered at 0 in $\Re^2$, and let $\Re^2 - B_r$ denote $\Re^2$ excluding this ball. If $r_1 > r_2$, let $B_{r_1} - B_{r_2}$ denote B_{r_1} excluding B_{r_2}. By Feller (1971, page 527) (or simply a short calculation), we have the first inequality in the following display, where the rest follow from (46) and our assumptions.

$$
\begin{aligned}
\sup_{ijt}\left|\psi_{ijt}^{X\varepsilon}(s_1, s_2)\right| \quad &\leq \quad \sup_{ijt}\left|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_t^\varepsilon(s_1 x_i + s_2 x_j)\, p_{ijt}^X(x_i, x_j)\, dx_i dx_j\right| \qquad (52) \\[2mm]
&\leq \quad \sup_{ijt}\left|\int_{\Re^2 - B_{|s|^{1/4}}}\psi_t^\varepsilon(s_1 x_i + s_2 x_j)\, p_{ijt}^X(x_i, x_j)\, dx\right| + \\[2mm]
&\qquad \sup_{ijt}\left|\int_{B_{1/|s|}}\psi_t^\varepsilon(s_1 x_i + s_2 x_j)\, p_{ijt}^X(x_i, x_j)\, dx\right| + \\[2mm]
&\qquad \sup_{ijt}\left|\int_{B_{|s|^{1/4}} - B_{1/|s|}}\psi_t^\varepsilon(s_1 x_i + s_2 x_j)\, p_{ijt}^X(x_i, x_j)\, dx\right| \\[2mm]
&\leq \quad \sup_{ijt}\int_{\Re^2 - B_{|s|^{1/4}}}|\psi_t^\varepsilon(s_1 x_i + s_2 x_j)|\, p_{ijt}^X(x_i, x_j)\, dx +
\end{aligned}
$$

$$\sup_{ijt} \int_{B_{1/|s|}} |\psi_t^\varepsilon (s_1 x_i + s_2 x_j)| \, p_{ijt}^X (x_i, x_j) \, dx +$$

$$\sup_{ijt} \int_{B_{|s|^{1/4} - B_{1/|s|}}} |\psi_t^\varepsilon (s_1 x_i + s_2 x_j)| \, p_{ijt}^X (x_i, x_j) \, dx$$

$$\leq \quad \sup_{ijt} \int_{\Re^2 - B_{|s|^{1/4}}} p_{ijt}^X (x_i, x_j) \, dx +$$

$$\sup_{ijt} \int_{B_{1/|s|}} p_{ijt}^X (x_i, x_j) \, dx +$$

$$\sup_{ijt} \int_A p_{ijt}^X (x_i, x_j) \, dx +$$

$$\sup_{ijt} \int_{B_{|s|^{1/4} - B_{1/|s|} - A}} |\psi_t^\varepsilon (s_1 x_i + s_2 x_j)| \, p_{ijt}^X (x_i, x_j) \, dx$$

$$\leq \quad \frac{C}{|s|^{1/2}} + \frac{C_2}{|s|} + 2 \, |s|^{1/2} \, \gamma +$$

$$\sup_{ijt} \int_{B_{|s|^{1/4} - B_{1/|s|} - A}} (s_1 x_i + s_2 x_j)^{-2} \, p_{ijt}^X (x_i, x_j) \, dx$$

$$= \quad \frac{C}{|s|^{1/2}} + \frac{C_2}{|s|} + 2 \, |s|^{1/2} \, \gamma +$$

$$\sup_{ijt} \int_{B_{|s|^{1/4} - B_{1/|s|} - A}} |s|^{-2} \, |x|^{-2} \cos^{-2} \theta_{s,x} p_{ijt}^X (x_i, x_j) \, dx$$

$$\leq \quad \frac{C}{|s|^{1/2}} + \frac{C_2}{|s|} + C_3 \, |s|^{1/2} \, \gamma +$$

$$\frac{\pi^2}{4} \, |s|^{-2} \, \gamma^{-2} \sup_{ijt} \int_{B_{|s|^{1/4} - B_{1/|s|} - A}} |x|^{-2} \, p_{ijt}^X (x_i, x_j) \, dx$$

$$\leq \quad \frac{C}{|s|^{1/2}} + \frac{C_2}{|s|} + C_3 \, |s|^{1/2} \, \gamma +$$

$$\frac{\pi^2}{4} \, |s|^{-2} \, \gamma^{-2} C_4 \int_{B_{|s|^{1/4}}} \frac{1}{r} dr d\theta$$

$$\leq \quad \frac{C}{|s|^{1/2}} + \frac{C_2}{|s|} + C_3 \, |s|^{1/2} \, \gamma +$$

$$\frac{\pi^2}{4} \, |s|^{-2} \, \gamma^{-2} C_4 2\pi \frac{1}{4} \ln (|s|)$$

$$\leq \quad \frac{C}{|s|^{1/2}} + \frac{C_2}{|s|} + \frac{C_3}{|s|^{1/4}} + \frac{C_5}{|s|^{1/2}} \ln (|s|)$$

$$\leq \quad \frac{C}{|s|^{1/2}} + \frac{C_2}{|s|} + \frac{C_3}{|s|^{1/4}} + \frac{C_5}{|s|^{1/3}}$$

where the second-to-last inequality follows from our choice of $\gamma = |s|^{-3/4}$. Thus, Condition A is satisfied with $\alpha = \frac{1}{4}$, and we are finished.

Lemma 3 (Rates for Additional Densities) *Under Assumptions 1, 2, and 3, $\exists\, C, K_0 < \infty$ s.t. $\forall\, K \geq K_0$,*

$$\sup_s \left| \bar{f}_K(s) - \phi(s) \right| \quad \leq \quad CK^{-\frac{1}{4}} \log K \qquad (a)$$

$$\sup_{i,\hat{b}_i} \left| m_{iK}\left(\hat{b}_i\right) - m_\phi\left(\hat{b}_i\right) \right| \quad \leq \quad CK^{-\frac{1}{4}} \log K \qquad (b)$$

$$\sup_{i,\hat{b}_i} \left| \bar{m}_K\left(\hat{b}_i\right) - m_\phi\left(\hat{b}_i\right) \right| \quad \leq \quad CK^{-\frac{1}{4}} \log K \qquad (c)$$

$$\sup_{i,\hat{b}_i} \left| m'_{iK}\left(\hat{b}_i\right) - m'_\phi\left(\hat{b}_i\right) \right| \quad \leq \quad CK^{-\frac{1}{8}} \log K \qquad (d)$$

$$\sup_{i,\hat{b}_i} \left| \bar{m}'_K\left(\hat{b}_i\right) - m'_\phi\left(\hat{b}_i\right) \right| \quad \leq \quad CK^{-\frac{1}{8}} \log K \qquad (e)$$

$$\sup_{i,j,\hat{b}_i,\hat{b}_j} \left| m_{ijK}\left(\hat{b}_i,\hat{b}_j\right) - m_\phi\left(\hat{b}_i\right) m_\phi\left(\hat{b}_j\right) \right| \quad \leq \quad CK^{-\frac{1}{6}} \log K \qquad (f)$$

$$\sup_{i,j,\hat{b}_i,\hat{b}_j} \left| \frac{\partial}{\partial \hat{b}_i} m_{ijK}\left(\hat{b}_i,\hat{b}_j\right) - m'_\phi\left(\hat{b}_i\right) m_\phi\left(\hat{b}_j\right) \right| \quad \leq \quad CK^{-\frac{1}{12}} \log K \qquad (g)$$

Proof: Part (a) follows from Lemma 1 and

$$\sup_s \left| \bar{f}_K(s) - \phi(s) \right| \qquad (53)$$

$$= \quad \sup_s \left| \frac{1}{K} \sum_{i=1}^K \left(f_{iK}(s) - \phi(s) \right) \right|$$

$$\leq \quad \frac{1}{K} \sum_{i=1}^K \sup_s \left| f_{iK}(s) - \phi(s) \right|$$

$$\leq \quad CK^{-1/4} \log K.$$

A-31

Part (b) follows from Lemma 1 and

$$\sup_{i,\hat{b}_i} \left| m_{iK}\left(\hat{b}_i\right) - m_\phi\left(\hat{b}_i\right) \right| \tag{54}$$

$$= \sup_{i,\hat{b}_i} \left| \int_{-\infty}^{\infty} \left[f_{iK}\left(\hat{b}_i - b_i\right) - \phi\left(\hat{b}_i - b_i\right) \right] dG\left(b_i\right) \right|$$

$$\leq \int_{-\infty}^{\infty} \sup_{i,\hat{b}_i} \left| f_{iK}\left(\hat{b}_i - b_i\right) - \phi\left(\hat{b}_i - b_i\right) \right| dG\left(b_i\right)$$

$$\leq CK^{-1/4}\log K$$

Part (c) follows from part (b) in exactly the same way that part (a) follows from Lemma 1. Part (d) follows from Lemma 1 and

$$\sup_{i,\hat{b}_i} \left| m'_{iK}\left(\hat{b}_i\right) - m'_\phi\left(\hat{b}_i\right) \right| \tag{55}$$

$$= \sup_{i,\hat{b}_i} \left| \frac{d}{d\hat{b}_i} \int_{-\infty}^{\infty} \left[f_{iK}\left(\hat{b}_i - b_i\right) - \phi\left(\hat{b}_i - b_i\right) \right] dG\left(b_i\right) \right|$$

$$= \sup_{i,\hat{b}_i} \left| \int_{-\infty}^{\infty} \left[f'_{iK}\left(\hat{b}_i - b_i\right) - \phi'\left(\hat{b}_i - b_i\right) \right] dG\left(b_i\right) \right|$$

$$\leq \int_{-\infty}^{\infty} \sup_{i,\hat{b}_i} \left| f'_{iK}\left(\hat{b}_i - b_i\right) - \phi'\left(\hat{b}_i - b_i\right) \right| dG\left(b_i\right)$$

$$\leq CK^{-1/8}\log K$$

where we can interchange differentiation and integration because of the uniformly bounded derivatives of both of the likelihood functions (see Appendix B for the proof that $\exists K_0 < \infty$ such that f'_{iK} is uniformly bounded for all $K \geq K_0$, under our assumptions). Part (e) follows from Part (d) just as Part (a) follows from Lemma 1.

Part (f) follows from Lemma 2 and

$$\sup_{i,\hat{b}_i,\hat{b}_j} \left| m_{ijK}\left(\hat{b}_i,\hat{b}_j\right) - m_\phi\left(\hat{b}_i,\hat{b}_j\right) \right| \tag{56}$$

$$= \sup_{i,\hat{b}_i,\hat{b}_j} \left| \int_{-\infty}^{\infty} \left[f_{ijK}\left(\hat{b}_i - b_i, \hat{b}_j - b_j\right) - \phi\left(\hat{b}_i - b_i\right)\phi\left(\hat{b}_j - b_j\right) \right] dG\left(b_i\right) dG\left(b_j\right) \right|$$

$$\leq \int_{-\infty}^{\infty} \sup_{i,\hat{b}_i,\hat{b}_j} \left| f_{ijK}\left(\hat{b}_i - b_i, \hat{b}_j - b_j\right) - \phi\left(\hat{b}_i - b_i\right)\phi\left(\hat{b}_j - b_j\right) \right| dG\left(b_i\right) dG\left(b_j\right)$$

$$\leq CK^{-1/6}\log K.$$

Part (g) follows from Lemma 2 and

$$\sup_{i,\hat{b}_i,\hat{b}_j} \left| \frac{\partial}{\partial \hat{b}_j} m_{ijK}\left(\hat{b}_i,\hat{b}_j\right) - \frac{\partial}{\partial \hat{b}_j} m_\phi\left(\hat{b}_i,\hat{b}_j\right) \right| \tag{57}$$

$$= \sup_{i,\hat{b}_i,\hat{b}_j} \left| \frac{\partial}{\partial \hat{b}_i} \int_{-\infty}^{\infty} \left[f_{iK}\left(\hat{b}_i - b_i, \hat{b}_j - b_j\right) - \phi\left(\hat{b}_i - b_i\right)\phi\left(\hat{b}_j - b_j\right) \right] dG\left(b_i\right) \right|$$

$$= \sup_{i,\hat{b}_i,\hat{b}_j} \left| \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \hat{b}_i} f_{iK}\left(\hat{b}_i - b_i\right) - \phi'\left(\hat{b}_i - b_i\right)\phi\left(\hat{b}_j - b_j\right) \right] dG\left(b_i\right) \right|$$

$$\leq \int_{-\infty}^{\infty} \sup_{i,\hat{b}_i,\hat{b}_j} \left| \frac{\partial}{\partial \hat{b}_i} f_{iK}\left(\hat{b}_i - b_i\right) - \phi'\left(\hat{b}_i - b_i\right)\phi\left(\hat{b}_j - b_j\right) \right| dG\left(b_i\right)$$

$$\leq CK^{-1/12}\log K$$

where we can interchange differentiation and integration because of the uniformly bounded derivatives of both of the likelihood functions (see Appendix B for the proof that $\exists K_0 < \infty$ such that $\frac{\partial}{\partial \hat{b}_i} f_{ijK}$ is uniformly bounded for all $K \geq K_0$, under our assumptions).

Lemma 4 (Information and Related Bounds) *Assumptions 1, 2, 3, 6, and 7*

imply that

$$\int_{-\infty}^{\infty} \frac{\left(m'_\phi\left(x\right)\right)^2}{m_\phi\left(x\right)} dx \quad < \quad \infty \tag{58}$$

$$\int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K\left(x\right)}{\bar{m}_K\left(x\right)+s_K}\right)^2 \bar{m}_K\left(x\right) dx \quad \rightarrow \quad \int_{-\infty}^{\infty} \frac{\left(m'_\phi\left(x\right)\right)^2}{m_\phi\left(x\right)} dx \tag{59}$$

$$\int_{-\infty}^{\infty} \left(\frac{m'_\phi\left(x\right)}{m_\phi\left(x\right)+s_K}\right)^2 m_\phi\left(x\right) dx \quad \rightarrow \quad \int_{-\infty}^{\infty} \frac{\left(m'_\phi\left(x\right)\right)^2}{m_\phi\left(x\right)} dx \tag{60}$$

$$\int_{-\infty}^{\infty} \left(\frac{m'_\phi\left(x\right)}{m_\phi\left(x\right)+s_K}\right)^2 \bar{m}_K\left(x\right) dx \quad \rightarrow \quad \int_{-\infty}^{\infty} \frac{\left(m'_\phi\left(x\right)\right)^2}{m_\phi\left(x\right)} dx \tag{61}$$

$$r_G\left(\hat{b}^{NB}, \phi_K\right) \quad < \quad \infty \tag{62}$$

Further, if in addition Assumption 5 holds,

$$\sup_{x,\,\theta\in\Theta} \left|m'_\phi\left(x;\theta\right)\right| \quad < \quad \infty \tag{63}$$

Proof: We shall prove the above lines in order, beginning with the inequality of expression (58).

$$
\begin{aligned}
\int_{-\infty}^{\infty} \frac{\left(m'_\phi\left(x\right)\right)^2}{m_\phi\left(x\right)} dx \quad &= \quad \int_{-\infty}^{\infty} \left(\frac{m'_\phi\left(x\right)}{m_\phi\left(x\right)}\right)^2 m_\phi\left(x\right) dx \\
&= \quad \int_{-\infty}^{\infty} \left(\frac{x-\hat{b}^{NB}\left(x\right)}{\sigma_\varepsilon^2}\right)^2 m_\phi\left(x\right) dx \\
&\leq \quad \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} x^2 m_\phi\left(x\right) dx \\
&\quad + \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} \left(\hat{b}^{NB}\left(x\right)\right)^2 m_\phi\left(x\right) dx \\
&= \quad \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} x^2 m_\phi\left(x\right) dx \\
&\quad + \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} \left(\frac{\int_{-\infty}^{\infty} b\phi\left(x-b\right) dG\left(b\right)}{\int_{-\infty}^{\infty} \phi\left(x-b\right) dG\left(b\right)}\right)^2 m_\phi\left(x\right) dx
\end{aligned}
\tag{64}
$$

A-34

$$\leq \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} x^2 m_\phi(x)\, dx$$

$$+ \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} \frac{\int_{-\infty}^{\infty} b^2 \phi(x-b)\, dG(b)}{\int_{-\infty}^{\infty} \phi(x-b)\, dG(b)} m_\phi(x)\, dx$$

$$= \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} x^2 m_\phi(x)\, dx$$

$$+ \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b^2 \phi(x-b)\, dG(b)\, dx$$

$$= \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} x^2 m_\phi(x)\, dx$$

$$+ \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} b^2 \left\{ \int_{-\infty}^{\infty} \phi(x-b)\, dx \right\} dG(b)$$

$$= \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} x^2 m_\phi(x)\, dx$$

$$+ \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} b^2\, dG(b)$$

$$< \quad \infty$$

where the first equality follows from trivial manipulation, the second equality is due to the fact that $\hat{b}^{NB}(x) = x - \sigma_\varepsilon^2 \frac{m_\phi'(x)}{m_\phi(x)}$ due to the normal likelihood, the first inequality comes from the fact that $(a - b)^2 \leq 2a^2 + 2b^2$, the third equality is by definition, the second inequality is due to the convexity of the squaring function and an almost-sure-in-x Jensen's inequality result for conditional expectations (since $\hat{b}^{NB}(x)$ is the conditional expectation of b given x) (see Billingsley [1995, page 449, equation (34.7)]; note that the distribution for x for which the result holds almost surely is precisely the distribution with density m_ϕ, so the stated inequality holds), the fourth equality holds since $m_\phi(x) = \int_{-\infty}^{\infty} \phi(x-b)\, dG(b)$, the fifth equality follows from the Tonelli-Fubini theorem, the sixth equality holds since $\int_{-\infty}^{\infty} \phi(x-b)\, dx = 1 \; \forall b \in \mathcal{R}$, and the

final inequality is a direct result of the moment bounds given in Assumptions 1 and 6.

Now we shall prove the convergence in (59).

$$\int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} \right)^2 \bar{m}_K(x) \, dx \; - \; \int_{-\infty}^{\infty} \frac{\left(m'_\phi(x) \right)^2}{m_\phi(x)} dx \tag{65}$$

$$= \int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} \right)^2 (\bar{m}_K(x) - m_\phi(x)) \, dx \qquad \text{(Term I)}$$

$$+ \int_{-\infty}^{\infty} \left[\left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} \right)^2 - \left(\frac{m'_\phi(x)}{m_\phi(x)} \right)^2 \right] m_\phi(x) \, dx \qquad \text{(Term II)}$$

follows by adding and subtracting $\int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x)+s_K} \right)^2 m_\phi(x) \, dx$. First we show that Term I converges to zero. Let $z_K \to \infty$ such that $s_K^{-2} z_K K^{-1/4} \log K \to 0$ and $s_K^{-2} z_K^{-2} \to 0$, which is certainly possible in light of Assumption 7. Then

$$\int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} \right)^2 (\bar{m}_K(x) - m_\phi(x)) \, dx \tag{66}$$

$$\leq \; C_0 s_K^{-2} \int_{-\infty}^{\infty} (\bar{m}_K(x) - m_\phi(x)) \, dx$$

$$= \; C_0 s_K^{-2} \left\{ \begin{array}{l} \int_{-z_K}^{z_K} (\bar{m}_K(x) - m_\phi(x)) \, dx \\[2mm] + \int_{-\infty}^{-z_K} (\bar{m}_K(x) - m_\phi(x)) \, dx \\[2mm] + \int_{z_K}^{\infty} (\bar{m}_K(x) - m_\phi(x)) \, dx \end{array} \right\}$$

$$\leq \; C_0 s_K^{-2} \left\{ \begin{array}{l} 2 z_K C K^{-1/4} \log K \\[2mm] + \int_{-\infty}^{-z_K} (\bar{m}_K(x) - m_\phi(x)) \, dx \\[2mm] + \int_{z_K}^{\infty} (\bar{m}_K(x) - m_\phi(x)) \, dx \end{array} \right\}$$

A-36

$$\leq \quad C_0 s_K^{-2} \left\{ \begin{array}{c} 2z_K CK^{-1/4} \log K \\[2ex] + C_2 z_K^{-2} \end{array} \right\}$$

$$\rightarrow \quad 0$$

where the first inequality follows from the positivity of $\bar{m}_K(x)$ and the boundedness of $\bar{m}'_K(x)$ (from Lemma 3(c)), the first equality is trivial, the second inequality follows by using Lemma 3(c) to get $\int_{-z_K}^{z_K} (\bar{m}_K(x) - m_\phi(x))\, dx \leq \int_{-z_K}^{z_K} CK^{-1/4} \log K dx = 2z_K CK^{-1/4} \log K$, the third inequality follows from Markov's inequality and the (uniformly in K) finite second moments (due to Assumptions 1 and 6) of the distributions with densities $\bar{m}_K$ and m_ϕ. Finally, the convergence to zero is by the construction of z_K.

Second we demonstrate that Term II converges to zero. Use the fact that $(a^2 - b^2) = (a - b)^2 + 2b(a - b)$ to obtain

$$\int_{-\infty}^{\infty} \left[\left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} \right)^2 - \left(\frac{m'_\phi(x)}{m_\phi(x)} \right)^2 \right] m_\phi(x)\, dx \tag{67}$$

$$= \quad \int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} - \frac{m'_\phi(x)}{m_\phi(x)} \right)^2 m_\phi(x)\, dx \qquad \text{(Term IIA)}$$

$$+2 \int_{-\infty}^{\infty} \left(\frac{m'_\phi(x)}{m_\phi(x)} \right) \left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} - \frac{m'_\phi(x)}{m_\phi(x)} \right) m_\phi(x)\, dx. \quad \text{(Term IIB)}$$

We deal with Term IIA, then Term IIB. Add and subtract $\frac{m'_\phi(x)}{m_\phi(x) + s_K}$ inside the square, then use the fact that $(a + b) \leq 2a^2 + 2b^2$, to obtain

$$\int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} - \frac{m'_\phi(x)}{m_\phi(x)} \right)^2 m_\phi(x)\, dx \tag{68}$$

A-37

$$\leq \quad 2 \int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} - \frac{m'_\phi(x)}{m_\phi(x) + s_K} \right)^2 m_\phi(x)\, dx \qquad \text{(Term IIAi)}$$

$$+ 2 \int_{-\infty}^{\infty} \left(\frac{m'_\phi(x)}{m_\phi(x) + s_K} - \frac{m'_\phi(x)}{m_\phi(x)} \right)^2 m_\phi(x)\, dx. \qquad \text{(Term IIAii)}$$

Now Term IIAi can be shown to converge to zero by the following argument, which

uses the fact that $\left(\frac{a}{b} - \frac{c}{d} \right)^2 = \left(\frac{a}{b}\left(\frac{d-b}{d}\right) + \frac{a-c}{d} \right)^2 \leq 2\left(\frac{a}{b}\right)^2 \left(\frac{d-b}{d}\right)^2 + 2\left(\frac{a-c}{d}\right)^2$ to obtain

the first inequality below:

$$2 \int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} - \frac{m'_\phi(x)}{m_\phi(x) + s_K} \right)^2 m_\phi(x)\, dx \qquad\qquad (69)$$

$$\leq \quad 4 \int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} \right)^2 \left(\frac{m_\phi(x) - \bar{m}_K(x)}{m_\phi(x) + s_K} \right)^2 m_\phi(x)\, dx$$

$$+ 4 \int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K(x) - m'_\phi(x)}{m_\phi(x) + s_K} \right)^2 m_\phi(x)\, dx$$

$$\leq \quad 4 \int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} \right)^2 \left(\frac{CK^{-1/4}\log K}{m_\phi(x) + s_K} \right)^2 m_\phi(x)\, dx$$

$$+ 4 \int_{-\infty}^{\infty} \left(\frac{CK^{-1/8}\log K}{m_\phi(x) + s_K} \right)^2 m_\phi(x)\, dx$$

$$\leq \quad 4 s_K^{-4} C_2 K^{-1/2} \log K \ + \ 4 s_K^{-2} C^2 K^{-1/4} \log K$$

$$\rightarrow \quad 0$$

where the second inequality follows from Lemma 3(c,e) and the third inequality follows

from the nonnegativity of $\bar{m}_K(x)$ and the boundedness (by Lemma 3(c)) of $\bar{m}'_K(x)$.

Finally, the convergence to zero is by rate bounds given in equation (37).

To see that Term IIAii converges to zero, note that $\left(\frac{m'_\phi(x)}{m_\phi(x) + s_K} - \frac{m'_\phi(x)}{m_\phi(x)} \right)^2$ clearly

converges to zero pointwise (since $s_K \to 0$). Thus, if there is an integrable (with re-

spect to the measure having density m_ϕ) dominating function for $\left(\frac{m'_\phi(x)}{m_\phi(x) + s_K} - \frac{m'_\phi(x)}{m_\phi(x)} \right)^2$,

then we may apply the Dominated Convergence Theorem and be finished. But since $(a - b)^2 \leq 2a^2 + 2b^2$, we have that

$$
\left(\frac{m'_\phi(x)}{m_\phi(x) + s_K} - \frac{m'_\phi(x)}{m_\phi(x)} \right)^2 \tag{70}
$$
$$
\leq \ 2 \left(\frac{m'_\phi(x)}{m_\phi(x) + s_K} \right)^2 + 2 \left(\frac{m'_\phi(x)}{m_\phi(x)} \right)^2
$$
$$
\leq \ 2 \left(\frac{m'_\phi(x)}{m_\phi(x)} \right)^2 + 2 \left(\frac{m'_\phi(x)}{m_\phi(x)} \right)^2
$$
$$
= \ 4 \left(\frac{m'_\phi(x)}{m_\phi(x)} \right)^2
$$

and we know from our above proof of the inequality (58) that $\int_{-\infty}^{\infty} \left(\frac{m'_\phi(x)}{m_\phi(x)} \right)^2 m_\phi(x)\, dx \ =$ $\int_{-\infty}^{\infty} \frac{(m'_\phi(x))^2}{m_\phi(x)} dx \ < \ \infty$. Thus, Term IIAii converges to zero.

We must still deal with Term IIB:

$$
2 \int_{-\infty}^{\infty} \left(\frac{m'_\phi(x)}{m_\phi(x)} \right) \left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} - \frac{m'_\phi(x)}{m_\phi(x)} \right) m_\phi(x)\, dx \tag{71}
$$
$$
\leq \ 2 \left(\int_{-\infty}^{\infty} \left(\frac{m'_\phi(x)}{m_\phi(x)} \right)^2 m_\phi(x)\, dx \right)^{1/2}
$$
$$
\times \left(\int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} - \frac{m'_\phi(x)}{m_\phi(x)} \right)^2 m_\phi(x)\, dx \right)^{1/2}
$$

and we see that the first factor is, by our above proof of (58), simply a finite constant, while the second factor is just the square root of Term IIA. But this means that Term IIB converges to zero, so we have demonstrate the convergence displayed in (59).

To show that the convergence (60) holds, simply note that $\left(\frac{m'_\phi(x)}{m_\phi(x)+s_K} \right)^2 \rightarrow \left(\frac{m'_\phi(x)}{m_\phi(x)} \right)^2$ pointwise certainly holds, and further that $\left(\frac{m'_\phi(x)}{m_\phi(x)+s_K} \right)^2 \ \leq \ \left(\frac{m'_\phi(x)}{m_\phi(x)} \right)^2$ and

$\int_{-\infty}^{\infty} \left(\frac{m'_\phi(x)}{m_\phi(x)} \right)^2 m_\phi(x)\, dx = \int_{-\infty}^{\infty} \frac{\left(m'_\phi(x) \right)^2}{m_\phi(x)}\, dx < \infty$, so we may apply the Dominated

Convergence Theorem to obtain the desired conclusion.

To demonstrate the convergence (61) simply add and subtract $\int_{-\infty}^{\infty} \left(\frac{m'_\phi(x)}{m_\phi(x)+s_K} \right)^2 m_\phi(x)\, dx$ to obtain

$$
\begin{aligned}
& \int_{-\infty}^{\infty} \left(\frac{m'_\phi(x)}{m_\phi(x)+s_K} \right)^2 \bar{m}_K(x)\, dx \;-\; \int_{-\infty}^{\infty} \frac{\left(m'_\phi(x) \right)^2}{m_\phi(x)}\, dx \qquad\qquad (72) \\
= \;& \int_{-\infty}^{\infty} \left(\frac{m'_\phi(x)}{m_\phi(x)+s_K} \right)^2 \left(\bar{m}_K(x) - m_\phi(x) \right) dx \\
& + \int_{-\infty}^{\infty} \left(\frac{m'_\phi(x)}{m_\phi(x)+s_K} \right)^2 m_\phi(x)\, dx \;-\; \int_{-\infty}^{\infty} \frac{\left(m'_\phi(x) \right)^2}{m_\phi(x)}\, dx \\
\to \;& 0
\end{aligned}
$$

where the convergence to zero follows because the first integral can be treated in exactly the same way that Term I above was, while the difference of the second and third integrals goes to zero by the convergence (60) shown immediately above.

The finiteness of the Bayes risk in the normal problem, as claimed in (62), can be shown as follows:

$$
\begin{aligned}
r_G\left(\hat{b}^{NB}, \phi_K \right) &= \rho \int \int \frac{1}{K} \sum_{i=1}^{K} \left(\hat{b}_i^{NB}\left(\hat{b}_i \right) - b_i \right)^2 \phi_K\left(\hat{b} - b \right) d\hat{b}\, dG(b) \qquad (73) \\
&= \rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\hat{b}_i^{NB}\left(\hat{b}_i \right) - b_i \right)^2 \phi\left(\hat{b}_i - b_i \right) d\hat{b}_i\, dG(b_i) \\
&= \rho \int \int \left(\hat{b}_1^{NB}\left(\hat{b}_1 \right) - b_1 \right)^2 \phi\left(\hat{b}_1 - b_1 \right) d\hat{b}_1\, dG(b_1) \\
&\leq \rho \int \int b_1^2 \phi\left(\hat{b}_1 - b_1 \right) d\hat{b}_1\, dG(b_1) \\
&= \rho \int b_1^2 \left\{ \int \phi\left(\hat{b}_1 - b_1 \right) d\hat{b}_1 \right\} dG(b_1)
\end{aligned}
$$

A-40

$$= \rho \int b_1^2 dG(b_1)$$

$$< \infty$$

where the first equality is by definition, the second equality is trivial, the third equality

follows because each term in the sum of the expression on the second line is identical,

and we may rename all of the variables we are integrating over b_1 and $\hat{b}_1$, then take

the average over K such identical terms, and the first inequality holds because $\hat{b}^{NB}$

is a Bayes decision rule, so it produces a Bayes risk no higher than that of any

other decision rule; in particular, it performs no worse, in Bayes-risk terms, than the

constant estimator 0. The fourth equality follows from the Tonelli-Fubini theorem.

Finally, the second inequality holds by Assumption 6.

To see that the finite-derivative claim (63) holds in the parametric case, use the

following argument:

$$
\begin{aligned}
\sup_{x,\,\theta\in\Theta} \left| m'_\phi(x;\theta) \right| &= \sup_{x,\,\theta\in\Theta} \left| \frac{d}{dx} \int \phi(x-b)\, dG(b;\theta) \right| \qquad (74) \\
&= \sup_{x,\,\theta\in\Theta} \left| \int \phi'(x-b)\, dG(b;\theta) \right| \\
&\leq \sup_{\theta\in\Theta} \int \sup_x |\phi'(x-b)|\, dG(b;\theta) \\
&= \sup_{\theta\in\Theta} \int C_0\, dG(b;\theta) \\
&= C_0
\end{aligned}
$$

where the first equality is by definition, the second equality is by the form of the

normal likelihood, which satisfies the conditions of Dudley (1999, Corollary A.12 on

page 394), the inequality is due to the convexity of the sup function, and the third

equality follows by the direct calculation that $\sup_x |\phi'(x-b)| = \frac{1}{\sigma_\varepsilon^2 \sqrt{2\pi}} e^{-1/2} \; \forall b$,

and by defining $C_0 = \frac{1}{\sigma_\varepsilon^2 \sqrt{2\pi}} e^{-1/2}$. The fourth equality is trivial, since C_0 is a

constant with respect to both b and θ.

Lemma 5 (Lower Bounds for Densities) *Under Assumptions 1, 2, 3, 6, and 7,*

$\exists C_1, C_2, K_0 < \infty$ *s.t.* $\forall K \geq K_0$

$$\inf_{x \in [-d_K, d_K]} m_\phi(x) \geq C_1 K^{-1/32} \tag{a}$$

$$\inf_{i,\, x \in [-d_K, d_K]} m_{iK}(x) \geq C_2 K^{-1/32} \tag{b}$$

$$\sup_{x \in [-d_K, d_K]} \left\{ \frac{1}{m_\phi(x)} \right\} \leq \frac{1}{C_1} K^{1/32} \tag{c}$$

$$\sup_{i,\, x \in [-d_K, d_K]} \left\{ \frac{1}{m_{iK}(x)} \right\} \leq \frac{1}{C_2} K^{1/32} \tag{d}$$

where d_K is defined in Definition 1 in this appendix.

Proof: We prove the parts of the lemma in order, from (a) to (d).

$$
\begin{aligned}
\inf_{x \in [-d_K, d_K]} m_\phi(x) &= \inf_{x \in [-d_K, d_K]} \int_{-\infty}^{\infty} \phi(x-u)\, dG(u) \\
&\geq \int_{-\infty}^{\infty} \left\{ \inf_{x \in [-d_K, d_K]} \phi(x-u) \right\} dG(u) \\
&= \int_{-\infty}^{\infty} \left\{ \inf_{x \in [-d_K, d_K]} \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{1}{2\sigma_\varepsilon^2}(x-u)^2} \right\} dG(u) \\
&\geq \int_{-d_K}^{d_K} \left\{ \inf_{x \in [-d_K, d_K]} \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{1}{2\sigma_\varepsilon^2}(x-u)^2} \right\} dG(u) \\
&\geq \int_{-d_K}^{d_K} \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{1}{2\sigma_\varepsilon^2} 4 d_K^2}\, dG(u)
\end{aligned}
$$

A-42

$$
\begin{aligned}
&= \quad \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{1}{2\sigma_\varepsilon^2} 4 d_K^2} \int_{-d_K}^{d_K} dG(u) \\
&\geq \quad \frac{1}{D\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_\varepsilon^2} 4 d_K^2\right) \int_{-d_K}^{d_K} dG(u) \\
&\geq \quad \frac{1}{D\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_\varepsilon^2} 4 d_K^2\right) \left[1 - \frac{C_3}{d_K^2}\right] \\
&\geq \quad C_1 \exp\left(-\frac{1}{2\sigma_\varepsilon^2} 4 d_K^2\right) \\
&= \quad C_1 \exp\left(-\frac{1}{2\sigma_\varepsilon^2} 4\sigma_\varepsilon^2 \frac{\log K}{64}\right) \\
&= \quad C_1 K^{-1/32}
\end{aligned}
$$

where the first equality is by definition, the first inequality holds because the infimum function is concave, the second equality is again by definition, the second inequality is due to the fact that the integrand is nonnegative, the third inequality follows because the infimum of the normal density over x and u both in $[-d_K, d_K]$ can be no smaller than if x and u were $2d_K$ apart, the third equality follows because the integrand in the previous line is not a function of u, the fourth inequality follows from the fact that $\sigma_\varepsilon^2 \leq D < \infty$ by Assumption 1, the fifth inequality is by the existence of a second moment of G (Assumption 6) and Markov's inequality, the sixth equality holds because, for K beyond some K_0 large enough, we can simply note that $1 - \frac{C_3}{d_K^2}$ is greater than some positive constant, the fourth equality is by the definition of d_K in Assumption 7, and the last equality follows by simple calculation.

To see that the inequality of part (b) holds, note that $\inf_{i,\, x \in [-d_K, d_K]} m_{iK}(x) \geq \inf_{x \in [-d_K, d_K]} m_\phi(x) - CK^{-1/4}\log K \geq C_1 K^{-1/32} - CK^{-1/4}\log K$, where the

A-43

first inequality comes from Lemma 3(b) and the second comes from part (a) of this lemma, which we just proved. Now, for K beyond some K_0 which is sufficiently large, we can simply absorb the $K^{-1/4} \log K$ term into the constant on the $K^{-1/32}$ term, because the latter goes to zero more slowly. Doing so, we obtain the inequality of part (b). To prove (c) and (d), simply invert the relations in (a) and (b).

Lemma 6 (Kernel MSE Rates) *Under Assumptions 1, 2, 3, 4, 6, and 7,* $\exists\, C, K_0 < \infty$ *s.t.* $\forall\, K \geq K_0$,

$$\sup_{i,|\hat{b}_i| \leq d_K} E\left\{ \left[\hat{m}_{iK}\left(\hat{b}_i\right) - \bar{m}_K\left(\hat{b}_i\right) \right]^2 \mid \hat{b}_i \right\} \tag{a}$$

$$\leq\ C \left(\begin{array}{c} \frac{1}{h_K(K-1)} + h_K^2 + K^{-13/48}\log^2 K \\[2mm] +\ 2K^{-13/96} h_K \log K\ +\ K^{-1/2}\log^2 K \end{array} \right)$$

$$\sup_{i,|\hat{b}_i| \leq d_K} E\left\{ \left[\hat{m}'_{iK}\left(\hat{b}_i\right) - \bar{m}'_K\left(\hat{b}_i\right) \right]^2 \mid \hat{b}_i \right\} \tag{b}$$

$$\leq\ C \left(\begin{array}{c} \frac{1}{h_K^3(K-1)} + h_K^2 + K^{-5/48}\log^2 K \\[2mm] +\ 2K^{-5/96} h_K \log K\ +\ K^{-1/4}\log^2 K \end{array} \right)$$

$$\sup_{i,|\hat{b}_i| \leq d_K} \Pr\left[\left|\breve{l}_i\left(\hat{b}_i\right)\right| > q_K \mid \hat{b}_i \right] \tag{c}$$

$$\rightarrow\quad 0$$

Proof: If we have $\hat{b}_i \in [-d_K, d_K]$ then

$$E\left\{ \left[\hat{m}_{iK}\left(\hat{b}_i\right) - \bar{m}_K\left(\hat{b}_i\right) \right]^2 \mid \hat{b}_i \right\} \tag{75}$$

$$=\quad Var\left(\hat{m}_{iK}\left(\hat{b}_i\right) \mid \hat{b}_i \right)$$

A-44

$$+ \left(E\left[\hat{m}_{iK}\left(\hat{b}_i \right) \mid \hat{b}_i \right] - \bar{m}_K\left(\hat{b}_i \right) \right)^2$$

$$\leq \quad Var\left(\hat{m}_{iK}\left(\hat{b}_i \right) \mid \hat{b}_i \right) \qquad\qquad\qquad\qquad \text{(Term I)}$$

$$+ 2\left(E\left[\hat{m}_{iK}\left(\hat{b}_i \right) \mid \hat{b}_i \right] - m_\phi\left(\hat{b}_i \right) \right)^2 \qquad\qquad \text{(Term II)}$$

$$+ 2\left(m_\phi\left(\hat{b}_i \right) - \bar{m}_K\left(\hat{b}_i \right) \right)^2 . \qquad\qquad\qquad \text{(Term III)}$$

First consider Term I. Now,

$$Var\left(\hat{m}_{iK}\left(\hat{b}_i \right) \mid \hat{b}_i \right) \qquad\qquad\qquad\qquad\qquad\qquad (76)$$

$$= \quad Var\left(\frac{1}{h_K\left(K-1 \right)} \sum_{j\neq i} w\left(\frac{\hat{b}_i - \hat{b}_j}{h_K} \right) \mid \hat{b}_i \right)$$

$$\leq \quad \frac{1}{h_K^2\left(K-1 \right)^2} \sum_{j\neq i} \sum_{n\neq i} \left| Cov\left(w\left(\frac{\hat{b}_i - \hat{b}_j}{h_K} \right), w\left(\frac{\hat{b}_i - \hat{b}_n}{h_K} \right) \mid \hat{b}_i \right) \right|$$

$$\leq \quad \frac{1}{h_K^2\left(K-1 \right)^2} \sum_{j\neq i} \sum_{n\neq i} \tau\left(|j-n| \right) \sqrt{E\left(w\left(\frac{\hat{b}_i - \hat{b}_j}{h_K} \right)^2 \mid \hat{b}_i \right)} \times$$

$$\sqrt{E\left(w\left(\frac{\hat{b}_i - \hat{b}_j}{h_K} \right)^2 \mid \hat{b}_i \right)}$$

$$\leq \quad \frac{C}{h_K\left(K-1 \right)}$$

where the first equality is by definition, the first inequality is familiar, and the second inequality holds by the following argument: first, note that the σ-fields generated by random variables which are linear combinations of the elements of the $T \times 1$ vectors $\underline{X}_i$ (where the weights in the linear combinations are fixed) are certainly sub-σ-fields of the σ-fields generated by the $\underline{X}_i$ vectors themselves (intuitively, we may lose information by taking linear combinations, but we will certainly never gain information by

A-45

doing so). Thus, if we recall the notation of Assumption 4, and in addition define $\mathcal{G}_a^c$

as the σ-field generated by the random variables $\left\{ \hat{b}_i = \frac{1}{\sqrt{T}} \underline{X}_i \cdot \varepsilon + b_i : a \leq i \leq c \right\}$,

we see that for any b and for any ε, $\mathcal{G}_a^c$ is a sub-σ-field of $\mathcal{F}_a^c$, so that for any ε and

for any b,

$$
\begin{aligned}
\tau(n) \;\geq\; & \sup_m \sup_{x \in \mathcal{F}_1^m, y \in \mathcal{F}_{m+n}^\infty} \left| Corr\left(x, \, y \mid \underline{X}_j, \varepsilon \right) \right| \\
\geq\; & \sup_m \sup_{x \in \mathcal{G}_1^m(\varepsilon), y \in \mathcal{G}_{m+n}^\infty(\varepsilon)} \left| Corr\left(x, \, y \mid \frac{1}{\sqrt{T}} \underline{X}_j \cdot \varepsilon + b_j \right) \right|
\end{aligned}
$$

but this means that the mixing coefficients $\tau(n)$ apply to the $\hat{b}_i$ as well as the $\underline{X}_i$,

since we can certainly use the following bound (letting the σ-fields generated by the

$\hat{b}_i$ be denoted $\mathcal{G}_a^c$ as before):

$$
\begin{aligned}
& \sup_m \sup_{x \in \mathcal{G}_1^m, y \in \mathcal{G}_{m+n}^\infty} \left| Corr\left(x, \, y \mid \hat{b}_j, \varepsilon \right) \right| \\
=\; & \sup_m \sup_{x \in \mathcal{G}_1^m, y \in \mathcal{G}_{m+n}^\infty} E\left[\left| Corr\left(x, \, y \mid \hat{b}_j, \varepsilon, b \right) \right| \right] \\
\leq\; & \sup_{\varepsilon, b} \sup_m \sup_{x \in \mathcal{G}_1^m, y \in \mathcal{G}_{m+n}^\infty} \left| Corr\left(x, \, y \mid \hat{b}_j, \varepsilon, b \right) \right| \\
\leq\; & \tau(n)
\end{aligned}
$$

which allows us to make use of Doukhan (1994, Theorem 3 (5) on page 9) to con-

clude that, if u and v are $\mathcal{G}_1^m$-measurable and $\mathcal{G}_{m+n}^\infty$-measurable random variables,

respectively, and if $E[u^2], \, E[v^2] < \infty$, then

$$
|Cov(u, v)| \;\leq\; \tau(n) \sqrt{E[u^2]} \sqrt{E[v^2]}.
$$

A-46

The final inequality in display (76) follows from a change of variables (so that the argument of the kernel is no longer scaled by h_K), the boundedness of the kernel, and the summability of the $\tau(n)$ (from Assumption 4).

Next, turn to Term II in display (75). Because $|a||b-c|+|ab-c^2| \geq |c||a-c|$,

$$
\sup_{i,x \in [-d_K,d_K], y \in \Re} m_\phi(x) \left| \bar{m}_{iK}^C(y|x) - m_\phi(y) \right|
$$

$$
= \sup_{i,x \in [-d_K,d_K], y \in \Re} m_\phi(x) \left| \frac{1}{(K-1)} \sum_{j \neq i} \left(m_{ijK}^C(y|x) - m_\phi(y) \right) \right|
$$

$$
\leq \sup_{i,x \in [-d_K,d_K], y \in \Re} \frac{1}{(K-1)} \sum_{j \neq i} m_\phi(x) \left| m_{ijK}^C(y|x) - m_\phi(y) \right|
$$

$$
\leq \frac{1}{(K-1)} \sum_{j \neq i} \sup_{i,x \in [-d_K,d_K], y \in \Re} m_\phi(x) \left| m_{ijK}^C(y|x) - m_\phi(y) \right|
$$

$$
\leq \frac{1}{(K-1)} \sum_{j \neq i} \sup_{i,x \in [-d_K,d_K], y \in \Re} \left\{ \begin{array}{l} m_{ijK}^C(y|x) \left| m_{iK}(x) - m_\phi(x) \right| \\ \\ + \left| m_{ijK}(y,x) - m_\phi(y) m_\phi(x) \right| \end{array} \right\}
$$

$$
\leq \frac{1}{(K-1)} \sum_{j \neq i} \left\{ \begin{array}{l} \sup_{i,x \in [-d_K,d_K], y \in \Re} m_{ijK}^C(y|x) \left| m_{iK}(x) - m_\phi(x) \right| \\ \\ + \sup_{i,x \in [-d_K,d_K], y \in \Re} \left| m_{ijK}(y,x) - m_\phi(y) m_\phi(x) \right| \end{array} \right\}
$$

$$
\leq \left\{ \frac{1}{(K-1)} \sum_{j \neq i} \sup_{i,x \in [-d_K,d_K], y \in \Re} m_{ijK}^C(y|x) \left| m_{iK}(x) - m_\phi(x) \right| \right\}
$$
$$
+ CK^{-1/6} \log K
$$

$$
\leq \left\{ \frac{1}{(K-1)} \sum_{j \neq i} \left[\sup_{i,x \in [-d_K,d_K], y \in \Re} m_{ijK}^C(y|x) \right] \left[\sup_{i,x \in [-d_K,d_K], y \in \Re} \left| m_{iK}(x) - m_\phi(x) \right| \right] \right\}
$$
$$
+ CK^{-1/6} \log K
$$

$$
\leq \left\{ \frac{1}{(K-1)} \sum_{j \neq i} \sup_{i,x \in [-d_K,d_K], y \in \Re} m_{ijK}^C(y|x) CK^{-1/4} \log K \right\}
$$
$$
+ CK^{-1/6} \log K
$$

A-47

$$\leq \quad \left\{ \frac{1}{(K-1)} \sum_{j \neq i} C^* K^{1/32} C K^{-1/4} \log K \right\}$$
$$+ CK^{-1/6} \log K$$

$$= \quad C^* K^{1/32} C K^{-1/4} \log K \; + \; CK^{-1/6} \log K$$

$$\leq \quad C_2^* K^{-1/6} \log K$$

where the first equality is by definition, the first inequality is by the convexity of the

absolute value function, the second inequality is by the convexity of the sup func-

tion, the third inequality is by the fact stated immediately above the display, the

fourth inequality is again by the convexity of the sup function, the fifth inequality

is by Lemma 3(f), the sixth inequality is again by the properties of the sup function

(the sup of the product may never be greater than the product of the sups when

the arguments are nonnegative), the seventh inequality follows from Lemma 3(b),

and the eighth inequality follows because $m_{ijK}^C(y|x) = \frac{m_{ijK}(y,x)}{m_{iK}(x)}$ by definition, and

the numerator of this fraction is bounded, since the variance of ϕ is bounded below,

so the density ϕ is bounded, so the density m_ϕ is bounded, and Lemma 3(f) im-

plies that then the numerator of the fraction is bounded. But the denominator is

bounded by $\frac{1}{C_2} K^{1/32}$, as we know from Lemma 5(d). The second equality follows

since the summand does not depend on j, and the final inequality is due to the fact

that $K^{(1/32)-(1/4)} \log K = K^{-7/32} \log K$ converges to zero more quickly than

$K^{-1/6} \log K$, so the $K^{-7/32} \log K$ term may be absorbed into the constant on the

$K^{-1/6} \log K$ term for sufficiently large K.

But from the above calculation, it rapidly follows that

$$\sup_{i,x\in[-d_K,d_K],y\in\Re} \left| \bar{m}_{iK}^C (y|x) - m_\phi (y) \right| \tag{77}$$

$$= \sup_{i,x\in[-d_K,d_K],y\in\Re} \frac{m_\phi (x)}{m_\phi (x)} \left| \bar{m}_{iK}^C (y|x) - m_\phi (y) \right|$$

$$\leq \left\{ \sup_{i,x\in[-d_K,d_K],y\in\Re} \frac{1}{m_\phi (x)} \right\}$$

$$\times \left\{ \sup_{i,x\in[-d_K,d_K],y\in\Re} m_\phi (x) \left| \bar{m}_{iK}^C (y|x) - m_\phi (y) \right| \right\}$$

$$\leq \frac{1}{C_1} K^{1/32} C_2^* K^{-1/6} \log K$$

$$= C_3^* K^{-13/96} \log K$$

where the first equality follows by multiplication by one, the first inequality is due to the fact that the product of the sups is always at least as large as the sup of the products (for nonnegative arguments), the second inequality follows from the above calculation and Lemma 5(c), and the second equality comes from a trivial computation. Now, the result immediately above implies that

$$\sup_{i,\,\hat{b}_i\in[-d_K,d_K]} \left| E\left[\hat{m}_{iK} \left(\hat{b}_i\right) \mid \hat{b}_i \right] - m_\phi \left(\hat{b}_i\right) \right| \tag{78}$$

$$= \sup_{i,\,\hat{b}_i\in[-d_K,d_K]} \left| \frac{\frac{1}{(K-1)} \sum_{j\neq i} \int_{-\infty}^{\infty} w(z) m_{ijK}^C \left(\hat{b}_i - h_K z|\hat{b}_i\right) dz}{-m_\phi \left(\hat{b}_i\right)} \right|$$

$$= \sup_{i,\,\hat{b}_i\in[-d_K,d_K]} \left| \int_{-\infty}^{\infty} w(z) \bar{m}_{iK}^C \left(\hat{b}_i - h_K z|\hat{b}_i\right) dz - m_\phi \left(\hat{b}_i\right) \right|$$

$$= \sup_{i,\,\hat{b}_i\in[-d_K,d_K]} \left| \int_{-\infty}^{\infty} w(z) \left[\bar{m}_{iK}^C \left(\hat{b}_i - h_K z|\hat{b}_i\right) - m_\phi \left(\hat{b}_i\right) \right] dz \right|$$

A-49

$$\leq \sup_{i,\, \hat{b}_i \in [-d_K, d_K]} \int_{-\infty}^{\infty} w\left(z\right) \left| \bar{m}_{iK}^C \left(\hat{b}_i - h_K z | \hat{b}_i \right) - m_\phi \left(\hat{b}_i \right) \right| dz$$

$$\leq \int_{-\infty}^{\infty} w\left(z\right) \sup_{i,\, \hat{b}_i \in [-d_K, d_K]} \left| \bar{m}_{iK}^C \left(\hat{b}_i - h_K z | \hat{b}_i \right) - m_\phi \left(\hat{b}_i \right) \right| dz$$

$$\leq \int_{-\infty}^{\infty} w\left(z\right) \sup_{i,\, \hat{b}_i \in [-d_K, d_K]} \left| \bar{m}_{iK}^C \left(\hat{b}_i - h_K z | \hat{b}_i \right) - m_\phi \left(\hat{b}_i - h_K z \right) \right| dz$$

$$+ \int_{-\infty}^{\infty} w\left(z\right) \sup_{\hat{b}_i \in [-d_K, d_K]} \left| m_\phi \left(\hat{b}_i - h_K z \right) - m_\phi \left(\hat{b}_i \right) \right| dz$$

$$\leq \int_{-\infty}^{\infty} w\left(z\right) \sup_{i, x \in [-d_K, d_K], y \in \Re} \left| \bar{m}_{iK}^C \left(y | x \right) - m_\phi \left(y \right) \right| dz$$

$$+ \int_{-\infty}^{\infty} w\left(z\right) C_4^* h_K z \, dz$$

$$\leq \int_{-\infty}^{\infty} w\left(z\right) C_3^* K^{-13/96} \log K \, dz \;+\; C_4^* h_K \int_{-\infty}^{\infty} z w\left(z\right) dz$$

$$\leq C_3^* K^{-13/96} \log K \;+\; C_5^* h_K$$

where the first equality is by definition, the second equality follows from the linearity of integration, the third equality is due to the fact that $m_\phi\left(x\right) = \int_{-\infty}^{\infty} m_\phi\left(x\right) w\left(z\right) dz$ as long as w is a probability density, the first inequality is by the convexity of the absolute value function, the second inequality is by the convexity of the sup function, the third inequality follows from adding and subtracting $m_\phi\left(\hat{b}_i - h_K z\right)$ and then applying the triangle inequality (and the convexity of the sup function again), the fourth inequality follows because m_ϕ is uniformly Lipschitz continuous (since ϕ has a uniformly bounded first derivative, and we can safely interchange the derivative and the integral), the fifth inequality follows from the calculation immediately above the current one, and the sixth inequality is due to the fact that the kernel w has a bounded second, and thus first, moment (as well as the fact that it integrates to one,

A-50

in the case of the first of the two terms).

Finally, consider Term III in display (75) – we know that

$$\sup_{i, \hat{b}_i} \left(\bar{m}_K \left(\hat{b}_i \right) - m_\phi \left(\hat{b}_i \right) \right)^2 \leq CK^{-\frac{1}{2}} \log^2 K \tag{79}$$

from Lemma 3(c) above.

Substituting the bounds (76), (78), and (79) into display (75) yields part (a) of this lemma.

In the case part (b) of this lemma, concerning the derivative of the density, everything is entirely similar, except that the rates slow somewhat, as we use the rates from Lemma 3 which pertain to the derivatives, rather than those which pertain to the densities themselves.

We now have only to demonstrate part (c) of the lemma.

$$\sup_{i, |\hat{b}_i| \leq d_K} \Pr \left[\left| \breve{l}_i \left(\hat{b}_i \right) \right| > q_K \mid \hat{b}_i \right] \tag{80}$$

$$\leq \sup_{i, |\hat{b}_i| \leq d_K} \Pr \left[\left| \hat{m}'_{iK} \left(\hat{b}_i \right) \right| > s_K q_K \mid \hat{b}_i \right]$$

$$\leq \sup_{i, |\hat{b}_i| \leq d_K} \left(\frac{1}{s_K q_K} \right)^2 E \left[\left(\hat{m}'_{iK} \left(\hat{b}_i \right) \right)^2 \mid \hat{b}_i \right]$$

$$\leq C_0 \left(\frac{1}{s_K q_K} \right)^2$$

$$\rightarrow 0$$

where the first inequality follows from the definition of $\breve{l} = \frac{\hat{m}'}{\hat{m} + s_K}$ and the nonnegativity of $\hat{m}$, the second inequality is a direct application of Markov's inequality, and the

A-51

third inequality follows from part (b) of this lemma and the uniform boundedness of

$\bar{m}_K$, which itself follows from Lemma 3(c) and the uniform boundedness of m_ϕ, since

by Assumption 1 there exists D_2 such that $\sigma_\varepsilon^2 \geq D_2 > 0$. Finally, the convergence

to zero is by Assumption 7 and the rate results in display 37. Q.E.D.

Proof of Theorem 3:

Let $\hat{b}_i^{INB} \equiv \hat{b}_i - \sigma_\varepsilon^2 \frac{m_\phi'(\hat{b};\theta)}{m_\phi(\hat{b};\theta)+s_K}$, where s_K is as in Assumption 7. This is the

infeasible simple empirical Bayes estimator based on the normal-likelihood marginal

$m_\phi(x)$. Now write:

$$r_G\left(\hat{b}^{PEB}, f_K\right) - r_G\left(\hat{b}^{NB}, \phi_K\right) \tag{81a}$$

$$= r_G\left(\hat{b}^{PEB}, f_K\right) - r_G\left(\hat{b}^{INB}, f_K\right) \tag{Term I}$$

$$+ r_G\left(\hat{b}^{INB}, f_K\right) - r_G\left(\hat{b}^{INB}, \phi_K\right) \tag{Term II}$$

$$+ r_G\left(\hat{b}^{INB}, \phi_K\right) - r_G\left(\hat{b}^{NB}, \phi_K\right) \tag{Term III}$$

We will show that Terms I, II, and III all converge to zero. First consider Term

III:

$$r_G\left(\hat{b}^{INB}, \phi_K\right) - r_G\left(\hat{b}^{NB}, \phi_K\right) \tag{82}$$

$$= \rho \int \int \frac{1}{K} \sum_{i=1}^K \left(\hat{b}_i^{INB}\left(\hat{b}_i\right) - b_i\right)^2 \phi_K\left(\hat{b} - b\right) d\hat{b} dG_K\left(b;\theta\right)$$

$$- \rho \int \int \frac{1}{K} \sum_{i=1}^K \left(\hat{b}_i^{NB}\left(\hat{b}_i\right) - b_i\right)^2 \phi_K\left(\hat{b} - b\right) d\hat{b} dG_K\left(b;\theta\right)$$

$$= \rho \frac{1}{K} \sum_{i=1}^K \int \int \left(\hat{b}_i^{INB}\left(\hat{b}_i\right) - b_i\right)^2 \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i dG\left(b_i;\theta\right)$$

A-52

$$- \rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\hat{b}_i^{NB} \left(\hat{b}_i \right) - b_i \right)^2 \phi \left(\hat{b}_i - b_i \right) d\hat{b}_i dG \left(b_i; \theta \right)$$

$$= \rho \int \int \left(\hat{b}_1^{INB} \left(\hat{b}_1 \right) - b_1 \right)^2 \phi \left(\hat{b}_1 - b_1 \right) d\hat{b}_1 dG \left(b_1; \theta \right)$$

$$- \rho \int \int \left(\hat{b}_1^{NB} \left(\hat{b}_1 \right) - b_1 \right)^2 \phi \left(\hat{b}_1 - b_1 \right) d\hat{b}_1 dG \left(b_1; \theta \right)$$

$$\rightarrow \quad 0$$

where the third equality holds because the integrals which make up the summands on lines 3 and 4 above are identical, and depend only on the ith rescaled least-squares coefficient, $\hat{b}_i$. Thus, we could change variables in each of the summand integrals to make them all notationally the same, then simply perform the resulting (trivial) sum to get the equality. The convergence to 0 follows since $\hat{b}_1^{INB} \rightarrow \hat{b}_1^{NB}$ pointwise, as $s_K \rightarrow 0$, so if there is a dominating function for $\left(\hat{b}_1^{INB} \left(\hat{b}_1 \right) - b_1 \right)^2$ which has a finite integral with respect to the measure $\phi \left(\hat{b}_1 - b_1 \right) d\hat{b}_1 dG \left(b_1; \theta \right)$, then we may apply the dominated convergence theorem and be finished. But $\left(\hat{b}_1^{INB} \left(\hat{b}_1 \right) - b_1 \right)^2 \leq 2 \left(b_1 - \hat{b}_1 \right)^2 + 2\sigma_\varepsilon^4 \left(\frac{m_\phi' \left(\hat{b}; \theta \right)}{m_\phi \left(\hat{b}; \theta \right) + s_K} \right)^2$, the first term of which is integrable by Assumption 1; the second term of this bound is also integrable, by Lemma 4 (since it is less, for each K, than $2\sigma_\varepsilon^4 \left(\frac{m_\phi' \left(\hat{b}; \theta \right)}{m_\phi \left(\hat{b}; \theta \right)} \right)^2$).

The negative of Term II is handled as follows:

$$r_G \left(\hat{b}^{INB}, \phi_K \right) - r_G \left(\hat{b}^{INB}, f_K \right) \tag{83}$$

$$= \rho \int \int \frac{1}{K} \sum_{i=1}^{K} \left(\hat{b}_i^{INB} \left(\hat{b}_i \right) - b_i \right)^2 \phi_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)$$

$$- \rho \int \int \frac{1}{K} \sum_{i=1}^{K} \left(\hat{b}_i^{INB} \left(\hat{b}_i \right) - b_i \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)$$

$$= \rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\hat{b}_1^{INB} \left(\hat{b}_i \right) - b_i \right)^2 \phi \left(\hat{b}_i - b_i \right) d\hat{b}_i dG \left(b_i; \theta \right)$$

$$- \rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\hat{b}_1^{INB} \left(\hat{b}_i \right) - b_i \right)^2 \bar{f}_{iK} \left(\hat{b}_i - b_i \right) d\hat{b}_i dG \left(b_i; \theta \right)$$

$$= \rho \int \int \left(\hat{b}_1^{INB} \left(\hat{b}_1 \right) - b_1 \right)^2 \phi \left(\hat{b}_1 - b_1 \right) d\hat{b}_1 dG \left(b_1; \theta \right)$$

$$- \rho \int \int \left(\hat{b}_1^{INB} \left(\hat{b}_1 \right) - b_1 \right)^2 \bar{f}_K \left(\hat{b}_1 - b_1 \right) d\hat{b}_1 dG \left(b_1; \theta \right)$$

$$\rightarrow \quad 0$$

where the first equality is by definition, the second follows from noting that the decision rule $\hat{b}_i^{INB}$, as a function of $\hat{b}_i$, is the same for every i, so that $\hat{b}_i^{INB} = \hat{b}_1^{INB}$ as functions of $\hat{b}_i$, the third equality follows from change of variables and the definition of $\bar{f}_K$, and the convergence to zero comes from: first, the fact that $\bar{f}_K$ converges to ϕ pointwise from Lemma 3, and second, the fact that $\left(\hat{b}_1^{INB} \left(\hat{b}_1 \right) - b_1 \right)^2 \bar{f}_K \leq 2 \left(\hat{b}_1 - b_1 \right)^2 \bar{f}_K + 2\sigma_\varepsilon^4 \left(\frac{m_\phi'(\hat{b}_1)}{m_\phi(\hat{b}_1) + s_K} \right)^2 \bar{f}_K$. The first term of this bound is integrable (for every K) by Assumption 1, while the second term is integrable (for every K) by Lemma 4, and both terms have integrals which converge to the integrals of their limits, so we may use the Dominated Convergence Theorem and be finished.

Finally, we show that Term I converges to zero. We will actually work with the negative of Term I.

$$r_G \left(\hat{b}^{INB}, f_K \right) - r_G \left(\hat{b}^{PEB}, f_K \right) \tag{84}$$

$$= \rho \int \int \frac{1}{K} \sum_{i=1}^{K} \left(\hat{b}_i^{INB} \left(\hat{b}_i \right) - b_i \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)$$

A-54

$$
- \rho \int \int \frac{1}{K} \sum_{i=1}^{K} \left(\hat{b}_i^{PEB} \left(\hat{b}_i \right) - b_i \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)
$$

$$
= \rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\left(\hat{b}_i^{INB} \left(\hat{b}_i \right) - b_i \right)^2 - \left(\hat{b}_i^{PEB} \left(\hat{b}_i \right) - b_i \right)^2 \right)
$$

$$
\times f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)
$$

$$
= \rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left[\begin{array}{c} 2 \left(\hat{b}_i^{INB} \left(\hat{b}_i \right) - b_i \right) \\ \times \left(\sigma_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \theta \right)}{m_\phi \left(\hat{b}_i; \theta \right) + s_K} - \hat{\sigma}_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} \right) \\ - \left(\sigma_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \theta \right)}{m_\phi \left(\hat{b}_i; \theta \right) + s_K} - \hat{\sigma}_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} \right)^2 \end{array} \right]
$$

$$
\times f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)
$$

$$
= \rho \frac{1}{K} \sum_{i=1}^{K} \int \int 2 \left(\hat{b}_i^{INB} \left(\hat{b}_i \right) - b_i \right)
$$

$$
\times \left(\sigma_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \theta \right)}{m_\phi \left(\hat{b}_i; \theta \right) + s_K} - \hat{\sigma}_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} \right)
$$

$$
\times f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)
$$

$$
- \rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\sigma_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \theta \right)}{m_\phi \left(\hat{b}_i; \theta \right) + s_K} - \hat{\sigma}_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} \right)^2
$$

$$
\times f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)
$$

$$
\leq \rho \frac{2}{K} \sum_{i=1}^{K} \left\{ \int \int \left(\hat{b}_i^{INB} \left(\hat{b}_i \right) - b_i \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right) \right\}^{1/2}
$$

$$
\times \left\{ \begin{array}{c} \int \int \left(\sigma_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \theta \right)}{m_\phi \left(\hat{b}_i; \theta \right) + s_K} - \hat{\sigma}_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} \right)^2 \\ \times f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right) \end{array} \right\}^{1/2}
$$

$$
- \rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\sigma_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \theta \right)}{m_\phi \left(\hat{b}_i; \theta \right) + s_K} - \hat{\sigma}_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} \right)^2
$$

$$
\times f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)
$$

$$\leq \quad 2\rho \left\{ \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\hat{b}_i^{INB} \left(\hat{b}_i \right) - b_i \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right) \right\}^{1/2}$$

$$\times \left\{ \begin{array}{c} \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\sigma_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \theta \right)}{m_\phi \left(\hat{b}_i; \theta \right) + s_K} - \hat{\sigma}_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} \right)^2 \\ \times f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right) \end{array} \right\}^{1/2}$$

$$- \rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\sigma_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \theta \right)}{m_\phi \left(\hat{b}_i; \theta \right) + s_K} - \hat{\sigma}_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} \right)^2$$

$$\times f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)$$

where the third equality comes from the fact that $a^2 - b^2 = 2a \left(a - b \right) - \left(a - b \right)^2$, the first inequality is an application of Hölder's inequality, and the second inequality comes from the Cauchy-Schwartz inequality. The final expression above shows that we need only demonstrate

$$\rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\begin{array}{c} \sigma_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \theta \right)}{m_\phi \left(\hat{b}_i; \theta \right) + s_K} \\ - \hat{\sigma}_\varepsilon^2 \frac{m_\phi' \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} \end{array} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right) \qquad (85)$$

$$\rightarrow \quad 0$$

since

$$\frac{1}{K} \sum_{i=1}^{K} \int \int \left(\hat{b}_i^{INB} \left(\hat{b}_i \right) - b_i \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right) \qquad (86)$$

$$= \quad r_G \left(\hat{b}^{INB}, f_K \right)$$

$$\rightarrow \quad r_G \left(\hat{b}^{NB}, \phi_K \right)$$

from the facts that Term II and Term III converge to zero as shown above.

A-56

Now,

$$\rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\sigma_\varepsilon^2 \frac{m'_\phi \left(\hat{b}_i; \theta \right)}{m_\phi \left(\hat{b}_i; \theta \right) + s_K} - \hat{\sigma}_\varepsilon^2 \frac{m'_\phi \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right) \quad (87)$$

$$\leq \quad 2\rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2 \right)^2 \left(\frac{m'_\phi \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)$$

$$+ 2\rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\frac{m'_\phi \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} - \frac{m'_\phi \left(\hat{b}_i; \theta \right)}{m_\phi \left(\hat{b}_i; \theta \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)$$

$$\leq \quad \frac{C_1}{K s_K} + 2\rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left(\frac{m'_\phi \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} - \frac{m'_\phi \left(\hat{b}_i; \theta \right)}{m_\phi \left(\hat{b}_i; \theta \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)$$

$$\leq \quad \frac{C_1}{K s_K}$$

$$+ 4\rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left\{ \begin{array}{c} \left(\frac{m'_\phi \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} \right)^2 \left(\frac{m_\phi \left(\hat{b}_i; \theta \right) - m_\phi \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \theta \right) + s_K} \right)^2 \\ + \left(\frac{m'_\phi \left(\hat{b}_i; \theta \right) - m'_\phi \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \theta \right) + s_K} \right)^2 \end{array} \right\} f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)$$

$$\leq \quad \frac{C_1}{K s_K^2}$$

$$+ 4\rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left\{ \begin{array}{c} \left(\frac{m'_\phi \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K} \right)^2 \left(\frac{C_2 \left\| \theta - \hat{\theta} \right\|}{s_K} \right)^2 \\ + \left(\frac{C_2 \left\| \theta - \hat{\theta} \right\|}{s_K} \right)^2 \end{array} \right\} f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)$$

$$\leq \quad \frac{C_1}{K s_K^2}$$

$$+ 4\rho \frac{1}{K s_K^4} \left[1 + \sup_{x, \theta} \left\{ m'_\phi \left(x, \theta \right)^2 \right\} \right] \int \int K \left\| \hat{\theta} - \theta \right\|^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K \left(b; \theta \right)$$

$$\leq \quad \frac{C_1}{K s_K^2} + \frac{C_3}{K s_k^4}$$

$$\to \quad 0$$

where the first inequality exploits the fact that $\left(a + b \right)^2 \leq 2a^2 + 2b^2$ and uses the

trick of adding and subtracting $\sigma_\varepsilon^2 \frac{m'_\phi \left(\hat{b}_i; \hat{\theta} \right)}{m_\phi \left(\hat{b}_i; \hat{\theta} \right) + s_K}$, the second inequality comes from The-

orem 1 and the fact that $\left(m'_\phi\left(\hat{b}_i;\hat{\theta}\right)\right)^2$ is bounded (using the properties of normal convolutions), the third inequality uses the fact that $\left(\frac{a}{b}-\frac{c}{d}\right)^2 = \left(\frac{a}{b}\left(\frac{d-b}{d}\right)+\frac{a-c}{d}\right)^2 \leq 2\left(\frac{a}{b}\right)^2\left(\frac{d-b}{d}\right)^2 + 2\left(\frac{a-c}{d}\right)^2$, the fourth inequality uses the Lipschitz continuity (uniform in i and θ) of m' and m, which is implied by the Lipschitz assumption made on $g\left(b;\theta\right)$, the fifth inequality follows by rearrangement, and the sixth inequality uses the facts that $\left(m'_\phi\left(x;\theta\right)\right)^2$ is uniformly bounded, as above, and that, by Assumption 5, $K\left\|\hat{\theta}-\theta\right\|^2$ has uniformly bounded expectation along the sequence. The convergence of the final terms follows from Assumption 7 and display (37).

Please note that the score used in $\hat{b}^{PEB}$ implicitly depends upon $\hat{\sigma}^2_\varepsilon$, a dependence which is suppressed above. Since σ^2_ε is bounded above and below by Assumption 1, and $\hat{\sigma}^2_\varepsilon$ is $\sqrt{K}$-consistent by Theorem 1, this does not affect our results.

Now we prove part (ii) of the theorem:

We proceed to show that $r_G\left(\hat{b}^{NB},\phi\right)$ is the "asymptotic minimax Bayes risk" in the sense that

$$\lim_{K\to\infty}\inf_{\tilde{b}}\left\{\sup_{f_K}r_G\left(\tilde{b},f_K\right)\right\} \;=\; r_G\left(\hat{b}^{NB},\phi\right) \tag{88}$$

where the supremum is taken over the set of likelihoods satisfying the assumptions of the theorem. This is actually straightforward. First,

$$\lim\inf_{K\to\infty}\inf_{\tilde{b}}\left\{\sup_{f_K}r_G\left(\tilde{b},f_K\right)\right\} \;\geq\; r_G\left(\hat{b}^{NB},\phi\right) \tag{89}$$

A-58

follows from $\sup_{f_K} r_G\left(\tilde{b}, f_K\right) \geq r_G\left(\tilde{b}, \phi\right) \ \forall \tilde{b}, \forall K$, so that $\forall K$,

$$\inf_{\tilde{b}}\left\{\sup_{f_K} r_G\left(\tilde{b}, f_K\right)\right\} \geq \inf_{\tilde{b}}\left\{r_G\left(\tilde{b}, \phi\right)\right\} \tag{90}$$

$$= r_G\left(\hat{b}^{NB}, \phi\right).$$

Further,

$$\inf_{\tilde{b}}\left\{\sup_{f_K} r_G\left(\tilde{b}, f_K\right)\right\} \leq \sup_{f_K} r_G\left(\hat{b}^{INB}, f_K\right) \tag{91}$$

so that

$$\lim_{K\to\infty}\sup\inf_{\tilde{b}}\left\{\sup_{f_K} r_G\left(\tilde{b}, f_K\right)\right\} \leq \lim_{K\to\infty}\sup\sup_{f_K} r_G\left(\hat{b}^{INB}, f_K\right) \tag{92}$$

$$= r_G\left(\hat{b}^{NB}, \phi\right).$$

The last equality is obviously the key to the entire result. It holds because each of the bounding constants in the proof that Terms II and III converge to zero in part (i) depends on the primitive bounding constants in the assumptions. Thus, as long as these primitive bounding constants are fixed, the bounds in the proof of part (i) hold uniformly for f_K satisfying the assumptions, and the final equality above follows. Q.E.D.

Proof of Theorem 4:

Let $\hat{b}_i^{ISEB} = \hat{b}_i - \sigma_\varepsilon^2 \frac{\bar{m}_K'(\hat{b}_i)}{\bar{m}_K(\hat{b}_i)+s_K}$; this is the infeasible simple empirical Bayes estimator based on the true average marginal $\bar{m}_K$. Now write

$$r_G\left(\hat{b}^{NSEB}, f_K\right) - r_G\left(\hat{b}^{NB}, \phi_K\right) \tag{93}$$

A-59

$$= \quad r_G\left(\hat{b}^{NSEB}, f_K\right) - r_G\left(\hat{b}^{ISEB}, f_K\right) \qquad \text{(Term I)}$$

$$+ r_G\left(\hat{b}^{ISEB}, f_K\right) - r_G\left(\hat{b}^{NB}, \phi_K\right) \qquad \text{(Term II)}$$

We will show that both Term I and Term II go to zero. Recall that $r_G\left(\hat{b}^{NB}, \phi_K\right) < \infty$ from Lemma 4. First, let us deal with the easier term, Term II:

$$r_G\left(\hat{b}^{ISEB}, f_K\right) - r_G\left(\hat{b}^{NB}, \phi_K\right) \qquad (94)$$

$$= \quad \rho \int \int \frac{1}{K} \sum_{i=1}^{K} \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 f_K\left(\hat{b} - b\right) d\hat{b} dG_K\left(b\right)$$

$$- \rho \int \int \frac{1}{K} \sum_{i=1}^{K} \left(\hat{b}_i^{NB}\left(\hat{b}_i\right) - b_i\right)^2 \phi_K\left(\hat{b} - b\right) d\hat{b} dG_K\left(b\right)$$

$$= \quad \frac{1}{K} \sum_{i=1}^{K} \rho \int \int \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 f_{iK}\left(\hat{b}_i - b_i\right) d\hat{b}_i dG\left(b_i\right)$$

$$- \frac{1}{K} \sum_{i=1}^{K} \rho \int \int \left(\hat{b}_i^{NB}\left(\hat{b}_i\right) - b_i\right)^2 \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i dG\left(b_i\right)$$

$$= \quad \frac{1}{K} \sum_{i=1}^{K} \rho \int \int \left(\hat{b}_1^{ISEB}\left(x\right) - y\right)^2 f_{iK}\left(x - y\right) dx dG\left(y\right)$$

$$- \frac{1}{K} \sum_{i=1}^{K} \rho \int \int \left(\hat{b}_1^{NB}\left(x\right) - y\right)^2 \phi\left(x - y\right) dx dG\left(y\right)$$

$$= \quad \rho \int \int \left(\hat{b}_1^{ISEB}\left(x\right) - y\right)^2 \bar{f}_K\left(x - y\right) dx dG\left(y\right)$$

$$- \rho \int \int \left(\hat{b}_1^{NB}\left(x\right) - y\right)^2 \phi\left(x - y\right) dx dG\left(y\right)$$

The second equality comes from the fact that both $\hat{b}_i^{NB}$ and $\hat{b}_i^{ISEB}$ depend only on $\hat{b}_i$ and not on $\hat{b}_{-i}$, while the third equality follows from the fact that the functional form of both $\hat{b}_i^{NB}$ and $\hat{b}_i^{ISEB}$ is the same for each i.

From the bounds in Lemma 3, and the definition of $\hat{b}_i^{ISEB}$, we see that the first integrand in the final line converges to the second pointwise. Thus, if we can produce a

dominating function for the first integrand, we may apply the Dominated Convergence Theorem and be finished. But

$$
\left(\hat{b}_1^{ISEB}(x) - y \right)^2 \bar{f}_K(x - y) \tag{95}
$$

$$
= \left((x - y) - \sigma_\varepsilon^2 \frac{\bar{m}_K'(x)}{\bar{m}_K(x) + s_K} \right)^2 \bar{f}_K(x - y)
$$

$$
\leq 2(x - y)^2 \bar{f}_K(x - y) + 2\sigma_\varepsilon^4 \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) + s_K} \right)^2 \bar{f}_K(x - y)
$$

The first term in this bound is evidently integrable by Assumption 1. The second term is integrable by Lemma 4, so we have shown that Term II converges to zero.

Before dealing with the more challenging Term I, it is worth noting that we will never encounter a "division by zero" problem, because of three facts: 1) the limit density $m_\phi(x)$ is positive everywhere (since it is a convolution of a normal density with the prior); 2) the approximation error given by Lemma 3 is less than s_K by construction (for large enough K, see Assumption 7); 3) the denominator of $\hat{b}_i^{ISEB}$ is thus always greater than m_ϕ, since we take $\bar{m}_K(x) > m_\phi(x) - s_K$ (for large enough K) and add s_K to both sides of the inequality, so that we never encounter difficulties.

Now we shall demonstrate that Term I converges to zero. Our proof is in the spirit of Bickel *et al.* (1993, p. 405 ff) and van der Vaart (1988, p. 169 ff), but we extend their approaches to handle cross-sectional dependence of the $\hat{b}_i$ and to deal with a nonconstant sequence of likelihoods.

A-61

Note that, from (2.8) and (2.9) of the body of this paper, we can write $\hat{l}_{iK}(x) = \hat{l}\left(\hat{b}_i \, ; \, \hat{b}_{-i}\right)$, where the function $\hat{l}(\cdot, \cdot)$ does not depend on i. This representation, and a similar one for $\check{l}$, are adopted here. Although $\hat{\sigma}_\varepsilon^2$ depends on the full data, rather than only $\hat{b}$, this dependence is suppressed for notational convenience; the treatment below does, however, account for the fact that $\hat{\sigma}_\varepsilon^2$ is a function of the full data (see display (100)). Now,

$$r_G\left(\hat{b}^{ISEB}, f_K\right) - r_G\left(\hat{b}^{NSEB}, f_K\right) \tag{96}$$

$$= \frac{1}{K}\sum_{i=1}^{K}\rho \int \int \left[\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 - \left(\hat{b}_i^{NSEB}\left(\hat{b}\right) - b_i\right)^2\right] f_K\left(\hat{b} - b\right) d\hat{b} dG_K(b)$$

$$= \frac{1}{K}\sum_{i=1}^{K}\rho \int \int \left[\begin{array}{c} \left(\hat{b}_i - b_i - \sigma_\varepsilon^2 \frac{\bar{m}_K'(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right)^2 - \\ \left(\hat{b}_i - b_i - \hat{\sigma}_\varepsilon^2 \hat{l}\left(\hat{b}_i; \hat{b}_{-i}\right)\right)^2 \end{array} \right] f_K\left(\hat{b} - b\right) d\hat{b} dG_K(b)$$

$$= \frac{1}{K}\sum_{i=1}^{K}\rho \int \int \left[\begin{array}{c} -\left(\hat{\sigma}_\varepsilon^2 \hat{l}\left(\hat{b}_i; \hat{b}_{-i}\right) - \sigma_\varepsilon^2 \frac{\bar{m}_K'(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right)^2 + \\ 2\left(\hat{b}_i - b_i - \sigma_\varepsilon^2 \frac{\bar{m}_K'(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right)\left(\hat{\sigma}_\varepsilon^2 \hat{l}\left(\hat{b}_i; \hat{b}_{-i}\right) - \sigma_\varepsilon^2 \frac{\bar{m}_K'(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right) \end{array} \right]$$
$$\times f_K\left(\hat{b} - b\right) d\hat{b} dG_K(b)$$

$$= -\frac{1}{K}\sum_{i=1}^{K}\rho \int \int \left(\hat{\sigma}_\varepsilon^2 \hat{l}\left(\hat{b}_i; \hat{b}_{-i}\right) - \sigma_\varepsilon^2 \frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K}\right)^2 f_K\left(\hat{b} - b\right) d\hat{b} dG_K(b)$$

$$+ \frac{1}{K}\sum_{i=1}^{K}\rho \int \int 2\left(\hat{b}_i - b_i - \sigma_\varepsilon^2 \frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K}\right) \times$$
$$\left(\hat{\sigma}_\varepsilon^2 \hat{l}\left(\hat{b}_i; \hat{b}_{-i}\right) - \sigma_\varepsilon^2 \frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K}\right) f_K\left(\hat{b} - b\right) d\hat{b} dG_K(b)$$

$$\leq -\frac{1}{K}\sum_{i=1}^{K}\rho \int \int \left(\hat{\sigma}_\varepsilon^2 \hat{l}\left(\hat{b}_i; \hat{b}_{-i}\right) - \sigma_\varepsilon^2 \frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K}\right)^2 f_K\left(\hat{b} - b\right) d\hat{b} dG_K(b)$$

A-62

$$
+ \frac{2}{K} \sum_{i=1}^{K} \rho \left[\begin{array}{l} \left(\int \int \left(\hat{b}_i - b_i - \sigma_\varepsilon^2 \frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K(b) \right)^{\frac{1}{2}} \times \\ \left(\int \int \left(\hat{\sigma}_\varepsilon^2 \hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \sigma_\varepsilon^2 \frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K(b) \right)^{\frac{1}{2}} \end{array} \right]
$$

$$
\leq -\frac{1}{K} \sum_{i=1}^{K} \rho \int \int \left(\hat{\sigma}_\varepsilon^2 \hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \sigma_\varepsilon^2 \frac{\bar{m}'_K \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K(b)
$$

$$
+ 2 \left\{ \frac{1}{K} \sum_{i=1}^{K} \rho \int \int \left(\hat{b}_i - b_i - \sigma_\varepsilon^2 \frac{\bar{m}'_K \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K(b) \right\}^{\frac{1}{2}} \times
$$

$$
\left\{ \frac{1}{K} \sum_{i=1}^{K} \rho \int \int \left(\hat{\sigma}_\varepsilon^2 \hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \sigma_\varepsilon^2 \frac{\bar{m}'_K \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K(b) \right\}^{\frac{1}{2}}
$$

In the above, the first and second equalities are by definition; the third is a consequence of the fact that $a^2 - b^2 = 2a(a-b) - (a-b)^2$; the fourth equality simply breaks out the two terms of the integrand; the first inequality is an application of Hölder's inequality, and the second is an application of the Cauchy-Schwartz inequality.

From the final bound given above, we see that if we can show that

$$
\frac{1}{K} \sum_{i=1}^{K} \rho \int \int \left(\hat{\sigma}_\varepsilon^2 \hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \sigma_\varepsilon^2 \frac{\bar{m}'_K \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K(b) \tag{97}
$$

$$
\rightarrow \quad 0
$$

then we will have demonstrated that Term I converges to zero, because

$$
\frac{1}{K} \sum_{i=1}^{K} \rho \int \int \left(\hat{b}_i - b_i - \sigma_\varepsilon^2 \frac{\bar{m}'_K \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K(b) \tag{98}
$$

$$
= r_G \left(\hat{b}^{ISEB}, f_K \right)
$$

$$\rightarrow \quad r_G\left(\hat{b}^{NB}, \phi_K\right)$$

$$< \quad \infty.$$

from the fact that Term II converges to zero, and, as noted above, $r_G\left(\hat{b}^{NB}, \phi_K\right) < \infty$.

Using $(a+b)^2 \leq 2a^2 + 2b^2$, and adding and subtracting $\sigma_\varepsilon^2 \hat{l}\left(\hat{b}_i \ ; \ \hat{b}_{-i}\right)$, we have

$$\frac{1}{K}\sum_{i=1}^{K}\rho \int \int \left(\hat{\sigma}_\varepsilon^2 \hat{l}\left(\hat{b}_i; \hat{b}_{-i}\right) - \sigma_\varepsilon^2 \frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K}\right)^2 \times \tag{99}$$
$$f_K\left(\hat{b} - b\right) d\hat{b} dG_K\left(b\right)$$

$$\leq \quad \frac{2}{K}\sum_{i=1}^{K}\rho \int \int \left(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\right)^2 \hat{l}^2\left(\hat{b}_i; \hat{b}_{-i}\right) f_K\left(\hat{b} - b\right) d\hat{b} dG_K\left(b\right) \quad \text{(Term A)}$$

$$+\frac{2\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\rho \int \int \left(\hat{l}\left(\hat{b}_i; \hat{b}_{-i}\right) - \frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K}\right)^2 \times \quad \text{(Term B)}$$
$$f_K\left(\hat{b} - b\right) d\hat{b} dG_K\left(b\right)$$

which separates the problem of nonparametric score estimation from the problem of estimating the residual variance. Now, Term A satisfies

$$\frac{2}{K}\sum_{i=1}^{K}\rho \int \int \left(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\right)^2 \hat{l}^2\left(\hat{b}_i; \hat{b}_{-i}\right) f_K\left(\hat{b} - b\right) d\hat{b} dG_K\left(b\right) \tag{100}$$

$$\leq \quad \frac{2\rho q_K^2}{K}\sum_{i=1}^{K}\int \int \left(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\right)^2 f_K\left(\hat{b} - b\right) d\hat{b} dG_K\left(b\right)$$

$$= \quad 2\rho q_K^2 \int \int \left(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\right)^2 f_K\left(\hat{b} - b\right) d\hat{b} dG_K\left(b\right)$$

$$= \quad 2\rho q_K^2 Var\left[\hat{\sigma}_\varepsilon^2\right]$$

$$\leq \quad 2\rho q_K^2 \frac{C}{K}$$

$$\rightarrow \quad 0$$

A-64

where the first inequality comes from the truncation of our estimator $\hat{l}$ of the score, and the second inequality comes from Theorem 1. The convergence of the final bound to zero is by construction: by display (37), $\frac{q_K^2}{K} \to 0$. Thus, Term A converges to zero.

Now consider Term B. Define $D_i \equiv \left\{ \hat{b} : \left| \hat{b}_i \right| \leq \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128} \log K} \text{ and } \left| \breve{l} \left(\hat{b}_i; \hat{b}_{-i} \right) \right| \leq q_K \right\}$, and let $E_{D_i} \left[(\cdot) \right] \equiv \int_{\hat{b} \in D_i} (\cdot) \, m_K \left(\hat{b} \right) d\hat{b}$ (so that the area of integration is restricted, but in a way which may differ for each i). Now,

$$\frac{2\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \rho \int \int \left(\hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} dG_K (b) \quad (101)$$

$$= \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int \left(\hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 m_K \left(\hat{b} \right) d\hat{b}$$

$$= \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int E \left[\left(\hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 \mid \hat{b}_i \right] m_{iK} \left(\hat{b}_i \right) d\hat{b}_i$$

$$\leq \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{-\infty}^{\infty} \left(\frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 m_{iK} \left(\hat{b}_i \right) \times \qquad\qquad \text{(Term Bi)}$$

$$\left(\Pr \left(\left| \breve{l} \left(\hat{b}_i; \hat{b}_{-i} \right) \right| > q_K \mid \hat{b}_i \right) + \Pr \left(\left| \hat{b}_i \right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128} \log K} \mid \hat{b}_i \right) \right) d\hat{b}_i$$

$$+ \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} E_{D_i} \left[\left(\hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 \right] \qquad\qquad \text{(Term Bii)}$$

where the first equality is by the Tonelli-Fubini Theorem, the second is by definition, and the inequality follows from the truncation of the estimated score function according to the definition of D_i.

Consider Term Bi first.

$$\frac{2\rho\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\int_{-\infty}^{\infty}\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\,+\,s_K}\right)^2 m_{iK}\left(\hat{b}_i\right)\times \qquad (102)$$

$$\left(\begin{array}{c}\Pr\left(\left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right|>q_K\mid\hat{b}_i\right)+\\[2mm]\Pr\left(\left|\hat{b}_i\right|>\sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128}\log K}\mid\hat{b}_i\right)\end{array}\right)d\hat{b}_i$$

$$=\;\;\frac{2\rho\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\int_{-\infty}^{\infty}\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\,+\,s_K}\right)^2 m_{iK}\left(\hat{b}_i\right)$$

$$\times\Pr\left(\left|\hat{b}_i\right|>\sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128}\log K}\mid\hat{b}_i\right)d\hat{b}_i$$

$$+\frac{2\rho\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\int_{-\infty}^{\infty}\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\,+\,s_K}\right)^2 m_{iK}\left(\hat{b}_i\right)\times$$

$$\Pr\left(\left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right|>q_K\mid\hat{b}_i\right)d\hat{b}_i$$

$$\leq\;\;\frac{2\rho\sigma_\varepsilon^4}{Ks_K^2}C\sum_{i=1}^{K}\Pr\left(\left|\hat{b}_i\right|>\sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128}\log K}\right)$$

$$+\,2\rho\sigma_\varepsilon^4\int_{|x|>d_K}\left(\frac{\bar{m}_K'\left(x\right)}{\bar{m}_K\left(x\right)\,+\,s_K}\right)^2\bar{m}_K\left(x\right)dx$$

$$+\sup_{i,\left|\hat{b}_i\right|\leq d_K}\Pr\left(\left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right|>q_K\mid\hat{b}_i\right)$$

$$\times 2\rho\sigma_\varepsilon^4\int_{-d_K}^{d_K}\left(\frac{\bar{m}_K'\left(x\right)}{\bar{m}_K\left(x\right)\,+\,s_K}\right)^2\bar{m}_K\left(x\right)dx$$

$$\leq\;\;\frac{2\rho\sigma_\varepsilon^4}{Ks_K^2}C\sum_{i=1}^{K}\left\{\begin{array}{c}\Pr\left(\left|\hat{b}_i\right|>\sqrt{\frac{\sigma_\varepsilon^2}{256}\log K}\right)\\[2mm]+\Pr\left(\hat{\sigma}_\varepsilon^2\leq\frac{1}{2}\sigma_\varepsilon^2\right)\end{array}\right\}$$

$$+\,2\rho\sigma_\varepsilon^4\int_{|x|>d_K}\left(\frac{\bar{m}_K'\left(x\right)}{\bar{m}_K\left(x\right)\,+\,s_K}\right)^2\bar{m}_K\left(x\right)dx$$

$$+\sup_{i,\left|\hat{b}_i\right|\leq d_K}\Pr\left(\left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right|>q_K\mid\hat{b}_i\right)$$

$$\times 2\rho\sigma_\varepsilon^4 \int_{-d_K}^{d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) + s_K} \right)^2 \bar{m}_K(x)\, dx$$

$$\leq \quad C_2 s_K^{-2} \left\{ \frac{1}{\log K} + \frac{1}{K} \right\}$$

$$+ 2\rho\sigma_\varepsilon^4 \int_{|x|>d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) + s_K} \right)^2 \bar{m}_K(x)\, dx$$

$$+ \sup_{i,|\hat{b}_i| \leq d_K} \Pr\left(\left| \breve{l}\left(\hat{b}_i; \hat{b}_{-i}\right) \right| > q_K \mid \hat{b}_i \right)$$

$$\times 2\rho\sigma_\varepsilon^4 \int_{-d_K}^{d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) + s_K} \right)^2 \bar{m}_K(x)\, dx$$

The equality is trivial; the first inequality follows from the boundedness of $(\bar{m}_K'(x))^2$ (by Lemma 3(c) and the boundedness of $\left(m_\phi'(x)\right)^2$), the second inequality follows by $\Pr(A) \leq \Pr(B) + \Pr(C)$ whenever $A \subset (B \cup C)$ (if $\left|\hat{b}_i\right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128} \log K}$, then either $\left|\hat{b}_i\right| > \sqrt{\frac{\sigma_\varepsilon^2}{256} \log K}$ or $\hat{\sigma}_\varepsilon^2 \leq \frac{1}{2}\sigma_\varepsilon^2$, or both), and the third inequality follows from Chebyshev's inequality and the variance bound of Theorem 1 (along with $\sigma_\varepsilon^2 > 0$). Of the terms in the final expression, we see immediately that the first term converges to zero by Assumption 7 and the second converges to zero by the uniform integrability of its integrand, which was shown in the course of the proof that Term II converges to zero, and by $d_K \to \infty$. The third term converges to zero by Lemmas 4 and 6(c). Thus, Term Bi converges to zero.

Finally, turn to Term Bii. Consider the ith term of the average which makes up Term Bii. We define $E_{D_i}^{out}[(\cdot)]$ to be $\int_{\hat{b}\in D_i \cap \{\hat{\sigma}_\varepsilon^2 > 2\sigma_\varepsilon^2\}} (\cdot)\, m_K\left(\hat{b}\right) d\hat{b}$ and $E_{D_i}^{cond}\left[(\cdot) \mid \hat{b}_i\right]$ to be $\int_{\hat{b}\in D_i} (\cdot)\, m_K\left(\hat{b}_{-i} \mid \hat{b}_i\right) d\hat{b}_{-i}$ (so that both the probability measure and the area of integration depend on $\hat{b}_i$). Since $\left(\frac{a}{b} - \frac{c}{d}\right)^2 = \left(\frac{a}{b}\left(\frac{d-b}{d}\right) + \frac{a-c}{d}\right)^2 \leq 2\left(\frac{a}{b}\right)^2\left(\frac{d-b}{d}\right)^2 +$

$2\left(\frac{a-c}{d}\right)^2$ We have:

$$E_{D_i}\left[\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right) - \frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K}\right)^2\right] \tag{103a}$$

$$\leq \quad 2E_{D_i}^{out}\left[\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right)^2\right] + 2E_{D_i}^{out}\left[\left(\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K}\right)^2\right]$$

$$+ \int_{-d_K}^{d_K} E_{D_i}^{cond}\left[\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right) - \frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K}\right)^2 |\hat{b}_i\right] m_{iK}\left(\hat{b}_i\right) d\hat{b}_i$$

$$\leq \quad 2C\left(q_K^2 + s_K^{-2}\right)\Pr\left(\hat{\sigma}_\varepsilon^2 > 2\sigma_\varepsilon^2\right)$$

$$+2\int_{-d_K}^{d_K}\left(\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K}\right)^2 E_{D_i}^{cond}\left[\left(\frac{\bar{m}_K\left(\hat{b}_i\right) - \hat{m}_{iK}\left(\hat{b}_i\right)}{\hat{m}_{iK}\left(\hat{b}_i\right) + s_K}\right)^2 |\hat{b}_i\right] m_{iK}\left(\hat{b}_i\right) d\hat{b}_i$$

$$+2\int_{-d_K}^{d_K} E_{D_i}^{cond}\left[\left(\frac{\hat{m}'_{iK}\left(\hat{b}_i\right) - \bar{m}'_K\left(\hat{b}_i\right)}{\hat{m}_{iK}\left(\hat{b}_i\right) + s_K}\right)^2 |\hat{b}_i\right] m_{iK}\left(\hat{b}_i\right) d\hat{b}_i$$

$$\leq \quad 2\left(q_K^2 + s_K^{-2}\right)\frac{C_2}{K}$$

$$+\frac{2}{s_K^2}\int_{-d_K}^{d_K}\left(\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K}\right)^2 \times \tag{Term Biia}$$

$$E_{D_i}\left[\left(\bar{m}_K\left(\hat{b}_i\right) - \hat{m}_{iK}\left(\hat{b}_i\right)\right)^2 |\hat{b}_i\right] m_{iK}\left(\hat{b}_i\right) d\hat{b}_i$$

$$+\frac{2}{s_K^2}\int_{-d_K}^{d_K} E_{D_i}\left[\left(\hat{m}'_{iK}\left(\hat{b}_i\right) - \bar{m}'_K\left(\hat{b}_i\right)\right)^2 |\hat{b}_i\right] m_{iK}\left(\hat{b}_i\right) d\hat{b}_i. \tag{Term Biib}$$

where the first inequality follows from splitting the area of integration and $2a^2 + 2b^2 \geq (a+b)^2$ (and from the definition of d_K), the second inequality is due to the truncation of our score estimator and the boundedness of $(\bar{m}'_K(x))^2$ (by Lemma 3(c) and the boundedness of $\left(m'_\phi(x)\right)^2$), and applying the fact that $\left(\frac{a}{b} - \frac{c}{d}\right)^2 = \left(\frac{a}{b}\left(\frac{d-b}{d}\right) + \frac{a-c}{d}\right)^2 \leq 2\left(\frac{a}{b}\right)^2\left(\frac{d-b}{d}\right)^2 + 2\left(\frac{a-c}{d}\right)^2$ to the third term in the previous line. The final inequality

is by Theorem 1 (for the first term) and is clear for the other two terms. The first term of the final line converges to zero (uniformly in i) due to Assumption 7, so we consider only the remaining two terms.

Term Biia satisfies

$$\frac{2}{s_K^2} \int_{-d_K}^{d_K} \left(\frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 \times \tag{104}$$

$$E_{D_i} \left[\left(\bar{m}_K \left(\hat{b}_i \right) - \hat{m}_{iK} \left(\hat{b}_i \right) \right)^2 | \hat{b}_i \right] m_{iK} \left(\hat{b}_i \right) d\hat{b}_i$$

$$\leq \frac{2}{s_K^2} C \left(\frac{1}{h_K \left(K - 1 \right)} + h_K^2 + K^{-13/48} \log^2 K + 2K^{-13/96} h_K \log K + K^{-1/2} \log^2 K \right)$$

$$\times \int_{-\infty}^{\infty} \left(\frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 m_{iK} \left(\hat{b}_i \right) d\hat{b}_i$$

by Lemma 6(a), since omitting the restriction to $\hat{b} \in D_i$ can only make the expectation larger. Also, Term Biib satisfies

$$\frac{2}{s_K^2} \int_{-d_K}^{d_K} E_{D_i} \left[\left(\hat{m}_{iK}' \left(\hat{b}_i \right) - \bar{m}_K' \left(\hat{b}_i \right) \right)^2 | \hat{b}_i, \hat{b} \in D_i \right] m_{iK} \left(\hat{b}_i \right) d\hat{b}_i$$

$$\leq \frac{2}{s_K^2} C \left(\frac{1}{h_K^3 \left(K - 1 \right)} + h_K^2 + K^{-5/48} \log^2 K + 2K^{-5/96} h_K \log K + K^{-1/4} \log^2 K \right)$$

by Lemma 6(b), by the same logic. Thus we have that

$$\text{Term Bii} \tag{105}$$

A-69

$$\leq \frac{4\rho\sigma_\varepsilon^4}{s_K^2}C\left(\begin{array}{c}\left(\begin{array}{c}\frac{1}{h_K(K-1)}+h_K^2+K^{-13/48}\log^2 K \\ +2K^{-13/96}h_K\log K + K^{-1/2}\log^2 K\end{array}\right) \\ \times\int_{-\infty}^{\infty}\left(\frac{\bar{m}_K'(\hat{b}_i)}{\bar{m}_K(\hat{b}_i)+s_K}\right)^2 m_{iK}\left(\hat{b}_i\right)d\hat{b}_i \\ +\left(\begin{array}{c}\frac{1}{h_K^3(K-1)}+h_K^2+K^{-5/48}\log^2 K \\ +2K^{-5/96}h_K\log K + K^{-1/4}\log^2 K\end{array}\right)\end{array}\right)$$

$$\rightarrow \quad 0$$

by display (37) and Lemma 4 (which applies upon averaging), and we are finished with the proof that $r_G\left(\hat{b}^{NSEB},f_K\right)-r_G\left(\hat{b}^{NB},\phi_K\right) \rightarrow 0$.

(ii) The proof of part (ii) exactly parallels the proof of Theorem 3, part (ii), with $\hat{b}^{ISEB}$ replacing $\hat{b}^{INB}$. Q.E.D.

Proof of Theorem 5:

(i) First, we show that permuting $\hat{b}_{-i}$ has no effect on the value of $\tilde{b}_i\left(\hat{b}\right)$ when $\tilde{b}\in\mathcal{B}$. Suppose we define $\mathbf{P}_{-i}$ to be some permutation which leaves the index i fixed but may permute the other indices in an arbitrary way. We will use the same notation for $\mathbf{P}_{-i}$ and for its restriction to $\hat{b}_{-i}$, since context should always make our meaning clear. Then, for any $i=1,\ldots,K$,

$$\tilde{b}_i\left(\mathbf{P}_{-i}\hat{b}\right) = \left(\mathbf{P}_{-i}\tilde{b}\right)_i\left(\hat{b}\right) \tag{106}$$

$$= \tilde{b}_i\left(\hat{b}\right)$$

where the first line follows directly from $\tilde{b}\left(\mathbf{P}\hat{b}\right) = \mathbf{P}\left(\tilde{b}\left(\hat{b}\right)\right)$ for any permutation $\mathbf{P}$ (the definition of equivariance) and the second line is by the definition of $\mathbf{P}_{-i}$. Since this is true for any such permutation $\mathbf{P}_{-i}$, we see that $\tilde{b}_i\left(\hat{b}\right)$ depends only on the values, not on the ordering, of the coordinates $\hat{b}_{-i}$. Thus, letting $\hat{G}_{K,-i}$ be the empirical c.d.f. of the (realized values) $\hat{b}_{-i}$, we may write

$$\tilde{b}_i\left(\hat{b}\right) \;=\; \tilde{b}_i\left(\hat{b}_i\;;\;\hat{G}_{K,-i}\right) \tag{107}$$

for any $i = 1, \ldots, K$.

Further, we have, for any $i = 1, \ldots, K$,

$$\begin{aligned}
\tilde{b}_i\left(\hat{b}\right) &= \tilde{b}_i\left(\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_{i-1}, \hat{b}_i, \hat{b}_{i+1}, \ldots, \hat{b}_K\right) \\
&= \tilde{b}_1\left(\hat{b}_i, \hat{b}_2, \ldots, \hat{b}_{i-1}, \hat{b}_1, \hat{b}_{i+1}, \ldots, \hat{b}_K\right) \\
&= \tilde{b}_1\left(\hat{b}_i\;;\;\hat{G}_{K,-i}\right)
\end{aligned} \tag{108}$$

where the first line is simply a rewriting for clarity, the second line follows by considering the permutation which transposes the indices i and 1 and applying the definition of equivariance, and the final equality is due to the fact shown in equation (107), setting $i = 1$ and noting that the equation shows that the argument treated specially is the one in the first coordinate position, which, here, is $\hat{b}_i$.

Now consider the ith component of the risk function when $\tilde{b} \in \mathcal{B}$,

$$\begin{aligned}
R_i\left(b,\,\tilde{b};\,\phi_K\right) &= E\left[\left(\tilde{b}_i\left(\hat{b}\right) - b_i\right)^2 \mid b\right] \\
&= E\left[\left(\tilde{b}_1\left(\hat{b}_i\;;\;\hat{G}_{K,-i}\right) - b_i\right)^2 \mid b\right]
\end{aligned} \tag{109}$$

A-71

where the first equality is by definition (with the conditioning on b stated for clarity), and the second equality is due to the fact that $\tilde{b}_i\left(\hat{b}\right) = \tilde{b}_1\left(\hat{b}_i \; ; \; \hat{G}_{K,-i}\right)$ as shown above. Note that $\hat{b}_i$ and $\hat{G}_{K,-i}$ are independent since, under ϕ_K, $\hat{b}_i$ and $\hat{b}_{-i}$ are independent, and $\hat{G}_{K,-i}$ is a function of $\hat{b}_{-i}$. Moreover, the only part of b which $\hat{b}_i$ depends on is b_i, and the only part of b that $\hat{G}_{K,-i}$ depends on is $\tilde{G}_{K,-i}$, the empirical c.d.f. of the elements of b_{-i} (observe that the fact that the $\hat{b}_j$ are i.i.d. is crucial in making this hold). Using these facts and the above, we may write

$$R_i\left(b, \; \tilde{b}; \; \phi_K\right) \;=\; E\left[\left(\tilde{b}_1\left(\hat{b}_i \; ; \; \hat{G}_{K,-i}\right) - b_i\right)^2 \;\big|\; b_i, \tilde{G}_{K,-i}\right]. \tag{110}$$

But this has the same functional form, as a function of b_i and $\tilde{G}_{K,-i}$, for every $i = 1, \ldots, K$, since the estimator is, for each i, the same function of $\hat{b}_i$ and $\hat{G}_{K,-i}$. Therefore we have that

$$
\begin{aligned}
R_i\left(b, \; \tilde{b}; \; \phi_K\right) \;&=\; E\left[\left(\tilde{b}_1\left(\hat{b}_i \; ; \; \hat{G}_{K,-i}\right) - b_i\right)^2 \;\big|\; b_i, \tilde{G}_{K,-i}\right] \tag{111}\\
&\equiv\; q\left(b_i \; ; \; \tilde{G}_{K,-i}\right) \\
&=\; q\left(b_i \; ; \; \tilde{G}_K\right)
\end{aligned}
$$

where $\tilde{G}_K$ is the empirical c.d.f. of the b_i (the last line is trivial) and we suppress the dependence on the estimator and the likelihood. This yields, for $\tilde{b} \in \mathcal{B}$,

$$
\begin{aligned}
R\left(b, \; \tilde{b}; \; \phi_K\right) \;&=\; \rho\frac{1}{K}\sum_{i=1}^{K} R_i\left(b, \; \tilde{b}; \; \phi_K\right) \tag{112}\\
&=\; \rho\frac{1}{K}\sum_{i=1}^{K} q\left(b_i \; ; \; \tilde{G}_K\right)
\end{aligned}
$$

$$
= \rho \int_{-\infty}^{\infty} q\left(z \; ; \; \tilde{G}_K\right) d\tilde{G}_K\left(z\right)
$$

$$
= \rho \int_{-\infty}^{\infty} E\left[\left(\tilde{b}_1\left(\hat{b}_{ind(z)}, \hat{G}_{K,-ind(z)}\right) - z\right)^2 \mid z, \tilde{G}_{K,-i}\right] d\tilde{G}_K\left(z\right)
$$

where z is simply a variable of integration, and the last line follows by definition, where $\hat{b}_{ind(z)}$ is just the coordinate of the least-squares coefficient vector corresponding to the coordinate of b whose value z takes on. Now, it is clear that the form of the estimator $\tilde{b}_1$ above is restricted in the way in which it may depend on $\hat{b}$. Since removing the restriction and seeking the best decision rule in an unrestricted way can only lower the risk, we have

$$
\inf_{\tilde{b} \in \mathcal{B}} R\left(b, \; \tilde{b}; \; \phi_K\right) \geq \inf_{b^*\left(\hat{b}\right)} \left\{ \rho \int_{-\infty}^{\infty} E\left[\left(b^*\left(\hat{b}\right) - z\right)^2 \mid z, \tilde{G}_{K,-i}\right] d\tilde{G}_K\left(z\right) \right\} \tag{113}
$$

and by standard Bayesian calculation, we obtain that the vector of coordinates

$$
\hat{b}_i^{NB}\left(\hat{b}, \tilde{G}_K\right) = \frac{\int_{-\infty}^{\infty} z\phi\left(\hat{b}_i - z\right) d\tilde{G}_K\left(z\right)}{\int_{-\infty}^{\infty} \phi\left(\hat{b}_i - z\right) d\tilde{G}_K\left(z\right)} \tag{114}
$$

is the solution to the problem

$$
\arg\inf_{b^*\left(\hat{b}\right)} \left\{ \rho \int_{-\infty}^{\infty} E\left[\left(b^*\left(\hat{b}\right) - z\right)^2 \mid z, \tilde{G}_{K,-i}\right] d\tilde{G}_K\left(z\right) \right\} \tag{115}
$$

so that we have

$$
\inf_{\tilde{b} \in \mathcal{B}} R\left(b, \; \tilde{b}; \; \phi_K\right) \geq R\left(b, \; \hat{b}^{NB}; \; \phi_K\right) = r_{\tilde{G}_K}\left(\hat{b}^{NB}, \phi\right). \tag{116}
$$

An important point to observe is that $\hat{b}^{NB}\left(\hat{b}, \tilde{G}_K\right)$ is not an estimator: it is infeasible, since it depends on information (that is, the empirical c.d.f. of the b_i) which is not part of the data.

A-73

(ii) We prove that $\left| R\left(b,\ \hat{b}^{NSEB};\ f_K\right) - R\left(b,\ \hat{b}^{NB}_{\tilde{G}_K};\ \phi_K\right) \right| \rightarrow 0$, then simply set $f_K = \phi_K$ and observe that our result is uniform in $\|b\|_2 \leq M$. Let $\hat{b}^{ISEB}_i = \hat{b}_i - \sigma^2_\varepsilon \frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}$; this is the infeasible simple empirical Bayes estimator based on the average marginal

$$\bar{m}_K(x) = \frac{1}{K} \sum_{j=1}^{K} \bar{f}_K(x - b_j)$$

(with respect to the empirical c. d. f. of the true b_j). Likewise, we have the modified definitions

$$m_\phi(x) = \frac{1}{K} \sum_{j=1}^{K} \phi(x - b_j)$$

$$m_{iK}(x) = \frac{1}{K} \sum_{j=1}^{K} f_{iK}(x - b_j).$$

With these definitions, and imposing $\|b\|_2 \leq M$ uniformly along the K sequence, we note that Lemmas 1 through 6 hold uniformly along the K sequence. This is a critical conclusion, and is used throughout the following. The reason we may make this observation is that the only feature of the prior used in the lemmas is that the variance is bounded, and our assumption that $\|b\|_2 \leq M$ implies that the (empirical) variance of the b_i is uniformly bounded along the K sequence.

Now write

$$R\left(b,\ \hat{b}^{NSEB};\ f_K\right) - R\left(b,\ \hat{b}^{NB}_{\tilde{G}_K};\ \phi_K\right) \tag{117}$$

$$= R\left(b,\ \hat{b}^{NSEB};\ f_K\right) - R\left(b,\ \hat{b}^{ISEB};\ f_K\right) \qquad \text{(Term I)}$$

$$+ R\left(b,\ \hat{b}^{ISEB};\ f_K\right) - R\left(b,\ \hat{b}^{NB}_{\tilde{G}_K};\ \phi_K\right) \qquad \text{(Term II)}$$

A-74

We will show that both Term I and Term II go to zero. Note that $\sup_{\|b\|_2 \leq M} \limsup_{K \to \infty} R\left(b,\ \hat{b}_{\tilde{G}_K}^{NB};\ \phi_K\right) < \infty$ (for any $M < \infty$, where $\|b\|^2 \equiv \frac{1}{K}\sum_{i=1}^{K} b_i^2$) from reasoning exactly similar to that of Lemma 4. First, let us deal with the easier term, Term II:

$$\left| R\left(b,\ \hat{b}^{ISEB};\ f_K\right) - R\left(b,\ \hat{b}_{\tilde{G}_K}^{NB};\ \phi_K\right) \right| \tag{118}$$

$$= \left| \begin{array}{c} \rho \int \frac{1}{K}\sum_{i=1}^{K}\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 f_K\left(\hat{b} - b\right) d\hat{b} \\ - \rho \int \frac{1}{K}\sum_{i=1}^{K}\left(\hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right) - b_i\right)^2 \phi_K\left(\hat{b} - b\right) d\hat{b} \end{array} \right|$$

$$= \left| \begin{array}{c} \frac{1}{K}\sum_{i=1}^{K} \rho \int \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 f_{iK}\left(\hat{b}_i - b_i\right) d\hat{b}_i \\ - \frac{1}{K}\sum_{i=1}^{K} \rho \int \left(\hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right) - b_i\right)^2 \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i \end{array} \right|$$

$$\leq \rho \frac{1}{K}\sum_{i=1}^{K} \int \left| \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 - \left(\hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right) - b_i\right)^2 \right|$$

$$\times \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i$$

$$+ \rho\frac{1}{K}\sum_{i=1}^{K} \int \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2$$

$$\times \left| f_{iK}\left(\hat{b}_i - b_i\right) - \phi\left(\hat{b}_i - b_i\right) \right| d\hat{b}_i$$

$$\leq \rho\frac{1}{K}\sum_{i=1}^{K} \int \left| \begin{array}{c} \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - \hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right)\right)^2 \\ + 2\left(\hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right) - b_i\right) \\ \times \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - \hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right)\right) \end{array} \right| \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i$$

$$+ \rho\frac{1}{K}\sum_{i=1}^{K} \int_{-z_K}^{z_K} \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 C K^{-1/4}\log K d\hat{b}_i$$

$$+ \rho\frac{1}{K}\sum_{i=1}^{K} \int_{|\hat{b}_i|>z_K} \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2$$

$$\times \left(f_{iK}\left(\hat{b}_i - b_i\right) + \phi\left(\hat{b}_i - b_i\right) \right) d\hat{b}_i$$

$$
\begin{aligned}
\leq \quad & \rho\frac{1}{K}\sum_{i=1}^{K}\int\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right)-\hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right)\right)^2\phi\left(\hat{b}_i-b_i\right)d\hat{b}_i\\
& +2\rho\frac{1}{K}\sum_{i=1}^{K}\left(\int\left(\hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right)-b_i\right)^2\phi\left(\hat{b}_i-b_i\right)d\hat{b}_i\right)^{1/2}\\
& \times\left(\int\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right)-\hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right)\right)^2\phi\left(\hat{b}_i-b_i\right)d\hat{b}_i\right)^{1/2}\\
& +\rho\frac{1}{K}\sum_{i=1}^{K}\int_{-z_K}^{z_K}\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right)-b_i\right)^2CK^{-1/4}\log K d\hat{b}_i\\
& +\rho\frac{1}{K}\sum_{i=1}^{K}\int_{|\hat{b}_i|>z_K}\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right)-b_i\right)^2\\
& \times\left(f_{iK}\left(\hat{b}_i-b_i\right)+\phi\left(\hat{b}_i-b_i\right)\right)d\hat{b}_i\\
\leq \quad & \rho\frac{1}{K}\sum_{i=1}^{K}\int\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right)-\hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right)\right)^2\phi\left(\hat{b}_i-b_i\right)d\hat{b}_i\\
& +2\rho\left(\frac{1}{K}\sum_{i=1}^{K}\int\left(\hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right)-b_i\right)^2\phi\left(\hat{b}_i-b_i\right)d\hat{b}_i\right)^{1/2}\\
& \times\left(\frac{1}{K}\sum_{i=1}^{K}\int\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right)-\hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right)\right)^2\phi\left(\hat{b}_i-b_i\right)d\hat{b}_i\right)^{1/2}\\
& +\rho\frac{1}{K}\sum_{i=1}^{K}\int_{-z_K}^{z_K}\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right)-b_i\right)^2CK^{-1/4}\log K d\hat{b}_i\\
& +\rho\frac{1}{K}\sum_{i=1}^{K}\int_{|\hat{b}_i|>z_K}\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right)-b_i\right)^2\\
& \times\left(f_{iK}\left(\hat{b}_i-b_i\right)+\phi\left(\hat{b}_i-b_i\right)\right)d\hat{b}_i\\
\\
\leq \quad & \rho\frac{\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\int\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)+s_K}-\frac{m_\phi'\left(\hat{b}_i\right)}{m_\phi\left(\hat{b}_i\right)}\right)^2\phi\left(\hat{b}_i-b_i\right)db_i\\
& +2\rho\left(\frac{1}{K}\sum_{i=1}^{K}\int\left(\hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right)-b_i\right)^2\phi\left(\hat{b}_i-b_i\right)d\hat{b}_i\right)^{1/2}\\
& \times\left(\frac{\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\int\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)+s_K}-\frac{m_\phi'\left(\hat{b}_i\right)}{m_\phi\left(\hat{b}_i\right)}\right)^2\phi\left(\hat{b}_i-b_i\right)d\hat{b}_i\right)^{1/2}
\end{aligned}
$$

A-76

$$+ \rho \frac{2}{K} \sum_{i=1}^{K} \int_{-z_K}^{z_K} \left(\hat{b}_i - b_i \right)^2 C K^{-1/4} \log K d\hat{b}_i$$

$$+ \rho \frac{2\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{-z_K}^{z_K} \left(\frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 C K^{-1/4} \log K d\hat{b}_i$$

$$+ \rho \frac{2\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{\left| \hat{b}_i \right| > z_K} \left(\frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2$$
$$\times \left(f_{iK} \left(\hat{b}_i - b_i \right) + \phi \left(\hat{b}_i - b_i \right) \right) d\hat{b}_i$$

$$+ \rho \frac{2\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{\left| \hat{b}_i \right| > z_K} \left(\hat{b}_i - b_i \right)^2$$
$$\times \left(f_{iK} \left(\hat{b}_i - b_i \right) + \phi \left(\hat{b}_i - b_i \right) \right) d\hat{b}_i$$

The second equality comes from the fact that both $\hat{b}_{\tilde{G}_K, i}^{NB}$ and $\hat{b}_i^{ISEB}$ depend only

on $\hat{b}_i$ and not on $\hat{b}_{-i}$, the first inequality is trivial, the second inequality follows from

$a^2 - b^2 = (a - b)^2 + 2b (a - b)$ applied to the integrand of the first term of the preceding

expression, splitting up the area of integration of the second term of the preceding

expression, and noting that $\left| f_{iK} \left(\hat{b}_i - b_i \right) - \phi \left(\hat{b}_i - b_i \right) \right| \leq f_{iK} \left(\hat{b}_i - b_i \right) + \phi \left(\hat{b}_i - b_i \right)$,

the third inequality follows from applying the triangle inequality and then Hölder's

inequality to the first term of the preceding display, the fourth inequality follows

from applying the Cauchy-Schwartz inequality to the second term of the preceding

display, the fifth inequality follows from rewriting the first two terms of the preceding

display and using $(a + b)^2 \leq 2a^2 + 2b^2$ on the last two terms of the preceding display,

and the last expression converges to zero by the Dominated Convergence Theorem

through Lemma 4 (for the first two terms), and $z_K = s_K^{-2}$ (for the remaining terms)

A-77

along with Chebyshev's inequality and the observation that $\left(\frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i)+s_K}\right)^2 \leq Cs_K^{-2}$

for sufficiently large K by Lemma 3(e), the boundedness of m'_ϕ, and the nonnegativity

of $\bar{m}_K$.

Now we shall demonstrate that Term I converges to zero. Our proof is in the spirit

of Bickel *et al.* (1993, p. 405 ff) and van der Vaart (1988, p. 169 ff), but we extend

their approaches to handle cross-sectional dependence of the $\hat{b}_i$ and to deal with a

nonconstant sequence of likelihoods.

Note that, from (2.8) and (2.9) of the body of this paper, we can write $\hat{l}_{iK}(x) =$

$\hat{l}\left(\hat{b}_i \ ; \ \hat{b}_{-i}\right)$, where the function $\hat{l}(\cdot, \cdot)$ does not depend on i. This representation, and

a similar representation for $\breve{l}$, are adopted here. Although $\hat{\sigma}_\varepsilon^2$ depends on the full

data, rather than only $\hat{b}$, this dependence is suppressed for notational convenience;

the treatment below does, however, account for the fact that $\hat{\sigma}_\varepsilon^2$ is a function of the

full data (see display (123)). Now,

$$R\left(b, \ \hat{b}^{ISEB}; \ f_K\right) - R\left(b, \ \hat{b}^{NSEB}; \ f_K\right) \tag{119}$$

$$= \frac{1}{K}\sum_{i=1}^{K}\rho\int\left[\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 - \left(\hat{b}_i^{NSEB}\left(\hat{b}\right) - b_i\right)^2\right]f_K\left(\hat{b} - b\right)d\hat{b}$$

$$= \frac{1}{K}\sum_{i=1}^{K}\rho\int\left[\begin{array}{c}\left(\hat{b}_i - b_i - \sigma_\varepsilon^2\frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i)+s_K}\right)^2 - \\ \left(\hat{b}_i - b_i - \hat{\sigma}_\varepsilon^2\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right)^2\end{array}\right]f_K\left(\hat{b} - b\right)d\hat{b}$$

$$= \frac{1}{K}\sum_{i=1}^{K}\rho\int\left[\begin{array}{c}-\left(\hat{\sigma}_\varepsilon^2\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right) - \sigma_\varepsilon^2\frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i)+s_K}\right)^2 + \\ 2\left(\hat{b}_i - b_i - \sigma_\varepsilon^2\frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i)+s_K}\right)\left(\hat{\sigma}_\varepsilon^2\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right) - \sigma_\varepsilon^2\frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i)+s_K}\right)\end{array}\right]$$

A-78

$$\times f_K \left(\hat{b} - b \right) d\hat{b}$$

$$= \quad -\frac{1}{K} \sum_{i=1}^{K} \rho \int \left(\hat{\sigma}_\varepsilon^2 \hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \sigma_\varepsilon^2 \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b}$$

$$+ \frac{1}{K} \sum_{i=1}^{K} \rho \int 2 \left(\hat{b}_i - b_i - \sigma_\varepsilon^2 \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right) \times$$

$$\left(\hat{\sigma}_\varepsilon^2 \hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \sigma_\varepsilon^2 \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right) f_K \left(\hat{b} - b \right) d\hat{b}$$

$$\leq \quad -\frac{1}{K} \sum_{i=1}^{K} \rho \int \left(\hat{\sigma}_\varepsilon^2 \hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \sigma_\varepsilon^2 \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b}$$

$$+ \frac{2}{K} \sum_{i=1}^{K} \rho \left[\begin{array}{c} \left(\int \left(\hat{b}_i - b_i - \sigma_\varepsilon^2 \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} \right)^{\frac{1}{2}} \times \\ \left(\int \left(\hat{\sigma}_\varepsilon^2 \hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \sigma_\varepsilon^2 \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} \right)^{\frac{1}{2}} \end{array} \right]$$

$$\leq \quad -\frac{1}{K} \sum_{i=1}^{K} \rho \int \left(\hat{\sigma}_\varepsilon^2 \hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \sigma_\varepsilon^2 \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b}$$

$$+ 2 \left\{ \frac{1}{K} \sum_{i=1}^{K} \rho \int \left(\hat{b}_i - b_i - \sigma_\varepsilon^2 \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} \right\}^{\frac{1}{2}} \times$$

$$\left\{ \frac{1}{K} \sum_{i=1}^{K} \rho \int \left(\hat{\sigma}_\varepsilon^2 \hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \sigma_\varepsilon^2 \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} \right\}^{\frac{1}{2}}$$

In the above, the first and second equalities are by definition; the third is a consequence of the fact that $a^2 - b^2 = 2a \left(a - b \right) - \left(a - b \right)^2$; the fourth equality simply breaks out the two terms of the integrand; the first inequality is an application of Hölder's inequality, and the second is an application of the Cauchy-Schwartz inequality.

From the final bound given above, we see that if we can show that

$$
\frac{1}{K} \sum_{i=1}^{K} \rho \int \left(\hat{\sigma}_{\varepsilon}^{2} \hat{l} \left(\hat{b}_{i}; \hat{b}_{-i} \right) - \sigma_{\varepsilon}^{2} \frac{\bar{m}_{K}' \left(\hat{b}_{i} \right)}{\bar{m}_{K} \left(\hat{b}_{i} \right) + s_{K}} \right)^{2} f_{K} \left(\hat{b} - b \right) d\hat{b} \qquad (120)
$$
$$
\rightarrow \quad 0
$$

then we will have demonstrated that Term I converges to zero, because

$$
\frac{1}{K} \sum_{i=1}^{K} \rho \int \left(\hat{b}_{i} - b_{i} - \sigma_{\varepsilon}^{2} \frac{\bar{m}_{K}' \left(\hat{b}_{i} \right)}{\bar{m}_{K} \left(\hat{b}_{i} \right) + s_{K}} \right)^{2} f_{K} \left(\hat{b} - b \right) d\hat{b} \qquad (121)
$$
$$
= \quad R \left(b, \, \hat{b}^{ISEB}; \, f_{K} \right)
$$
$$
< \quad \infty \text{ uniformly along the } K, \, T \text{ sequence}
$$

from the fact that Term II $= R \left(b, \, \hat{b}^{ISEB}; \, f_{K} \right) - R \left(b, \, \hat{b}_{\check{G}_{K}}^{NB}; \, \phi_{K} \right)$ converges to zero,

and, as noted above, $\sup_{\|b\| \leq M} \limsup_{K \to \infty} R \left(b, \, \hat{b}_{\check{G}_{K}}^{NB}; \, \phi_{K} \right) < \infty$.

Using $\left(a + b \right)^{2} \leq 2a^{2} + 2b^{2}$, and adding and subtracting $\sigma_{\varepsilon}^{2} \hat{l} \left(\hat{b}_{i}; \, \hat{b}_{-i} \right)$, we have

$$
\frac{1}{K} \sum_{i=1}^{K} \rho \int \left(\hat{\sigma}_{\varepsilon}^{2} \hat{l} \left(\hat{b}_{i}; \hat{b}_{-i} \right) - \sigma_{\varepsilon}^{2} \frac{\bar{m}_{K}' \left(\hat{b}_{i} \right)}{\bar{m}_{K} \left(\hat{b}_{i} \right) + s_{K}} \right)^{2} \times \qquad (122)
$$
$$
f_{K} \left(\hat{b} - b \right) d\hat{b}
$$
$$
\leq \quad \frac{2}{K} \sum_{i=1}^{K} \rho \int \left(\hat{\sigma}_{\varepsilon}^{2} - \sigma_{\varepsilon}^{2} \right)^{2} \hat{l}^{2} \left(\hat{b}_{i}; \hat{b}_{-i} \right) f_{K} \left(\hat{b} - b \right) d\hat{b} \qquad \text{(Term A)}
$$
$$
+ \frac{2\sigma_{\varepsilon}^{4}}{K} \sum_{i=1}^{K} \rho \int \left(\hat{l} \left(\hat{b}_{i}; \hat{b}_{-i} \right) - \frac{\bar{m}_{K}' \left(\hat{b}_{i} \right)}{\bar{m}_{K} \left(\hat{b}_{i} \right) + s_{K}} \right)^{2} \times \qquad \text{(Term B)}
$$
$$
f_{K} \left(\hat{b} - b \right) d\hat{b}
$$

A-80

which separates the problem of nonparametric score estimation from the problem of estimating the residual variance. Now, Term A satisfies

$$\frac{2}{K} \sum_{i=1}^{K} \rho \int \left(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2 \right)^2 \hat{l}^2 \left(\hat{b}_i; \hat{b}_{-i} \right) f_K \left(\hat{b} - b \right) d\hat{b} \tag{123}$$

$$\leq \frac{2\rho q_K^2}{K} \sum_{i=1}^{K} \int \left(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2 \right)^2 f_K \left(\hat{b} - b \right) d\hat{b}$$

$$= 2\rho q_K^2 \int \left(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2 \right)^2 f_K \left(\hat{b} - b \right) d\hat{b}$$

$$= 2\rho q_K^2 Var \left[\hat{\sigma}_\varepsilon^2 \right]$$

$$\leq 2\rho q_K^2 \frac{C}{K}$$

$$\to 0$$

where the first inequality comes from the truncation of our estimator $\hat{l}$ of the score, and the second inequality comes from Theorem 1. The convergence of the final bound to zero is by construction: by display (37), $\frac{q_K^2}{K} \to 0$. Thus, Term A converges to zero.

Now consider Term B. Define $D_i \equiv \left\{ \hat{b} : \left| \hat{b}_i \right| \leq \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128} \log K} \text{ and } \left| \breve{l} \left(\hat{b}_i; \hat{b}_{-i} \right) \right| \leq q_K \right\}$, and let $E_{D_i} \left[(\cdot) \right] \equiv \int_{\hat{b} \in D_i} (\cdot) f_K \left(\hat{b} - b \right) d\hat{b}$ (so that the area of integration is restricted, but in a way which may differ for each i). Now,

$$\frac{2\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \rho \int \left(\hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 f_K \left(\hat{b} - b \right) d\hat{b} \tag{124}$$

$$= \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int E \left[\left(\hat{l} \left(\hat{b}_i; \hat{b}_{-i} \right) - \frac{\bar{m}_K' \left(\hat{b}_i \right)}{\bar{m}_K \left(\hat{b}_i \right) + s_K} \right)^2 \mid \hat{b}_i, b \right] f_{iK} \left(\hat{b}_i - b_i \right) d\hat{b}_i$$

$$\leq \ \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K} \right)^2 f_{iK}\left(\hat{b}_i - b_i\right) \times \qquad \text{(Term Bi)}$$

$$\left(\Pr\left(\left| \check{l}\left(\hat{b}_i; \hat{b}_{-i}\right) \right| > q_K \mid \hat{b}_i, b \right) + \Pr\left(\left| \hat{b}_i \right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128} \log K} \mid \hat{b}_i, b \right) \right) d\hat{b}_i$$

$$+ \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} E_{D_i}\left[\left(\hat{l}\left(\hat{b}_i; \hat{b}_{-i}\right) - \frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K} \right)^2 \mid b \right] \qquad \text{(Term Bii)}$$

where the first equality is by definition, and the inequality follows from the truncation

of the estimated score function according to the definition of D_i.

Consider Term Bi first.

$$\frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K} \right)^2 f_{iK}\left(\hat{b}_i - b_i\right) \times \qquad (125)$$

$$\left(\begin{array}{c} \Pr\left(\left| \check{l}\left(\hat{b}_i; \hat{b}_{-i}\right) \right| > q_K \mid \hat{b}_i, b \right) + \\[2mm] \Pr\left(\left| \hat{b}_i \right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128} \log K} \mid \hat{b}_i, b \right) \end{array} \right) d\hat{b}_i$$

$$= \ \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K} \right)^2 f_{iK}\left(\hat{b}_i - b_i\right)$$

$$\times \Pr\left(\left| \hat{b}_i \right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128} \log K} \mid \hat{b}_i, b \right) d\hat{b}_i$$

$$+ \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K} \right)^2 f_{iK}\left(\hat{b}_i - b_i\right) \times$$

$$\Pr\left(\left| \check{l}\left(\hat{b}_i; \hat{b}_{-i}\right) \right| > q_K \mid \hat{b}_i, b \right) d\hat{b}_i$$

$$\leq \ \frac{2\rho\sigma_\varepsilon^4}{K s_K^2} C \sum_{i=1}^{K} \Pr\left(\left| \hat{b}_i \right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128} \log K} \mid b \right)$$

$$+ \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{|x|>d_K} \left(\frac{\bar{m}'_K\left(x\right)}{\bar{m}_K\left(x\right)\ +\ s_K} \right)^2 f_{iK}\left(x - b_i\right) dx$$

$$+ \sup_{i, \left| \hat{b}_i \right| \leq d_K} \Pr\left(\left| \check{l}\left(\hat{b}_i; \hat{b}_{-i}\right) \right| > q_K \mid \hat{b}_i, b \right)$$

A-82

$$\times \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{-d_K}^{d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) + s_K} \right)^2 f_{iK}(x - b_i)\, dx$$

$$\leq \frac{2\rho\sigma_\varepsilon^4}{K s_K^2} C \sum_{i=1}^{K} \left\{ \begin{array}{c} \Pr\left(\left| \hat{b}_i \right| > \sqrt{\frac{\sigma_\varepsilon^2}{256} \log K} \right) \\[2mm] + \Pr\left(\hat{\sigma}_\varepsilon^2 > \tfrac{1}{2}\sigma_\varepsilon^2 \right) \end{array} \right\}$$

$$+ \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{|x|>d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) + s_K} \right)^2 f_{iK}(x - b_i)\, dx$$

$$+ \sup_{i,\, \left| \hat{b}_i \right| \leq d_K} \Pr\left(\left| \breve{l}\left(\hat{b}_i; \hat{b}_{-i} \right) \right| > q_K \mid \hat{b}_i, b \right)$$

$$\times \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{-d_K}^{d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) + s_K} \right)^2 f_{iK}(x - b_i)\, dx$$

$$\leq C_2 s_K^{-2} \left\{ \frac{1}{\log K} + \frac{1}{K} \right\}$$

$$+ \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{|x|>d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) + s_K} \right)^2 f_{iK}(x - b_i)\, dx$$

$$+ \sup_{i,\, \left| \hat{b}_i \right| \leq d_K} \Pr\left(\left| \breve{l}\left(\hat{b}_i; \hat{b}_{-i} \right) \right| > q_K \mid \hat{b}_i, b \right)$$

$$\times \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{-d_K}^{d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) + s_K} \right)^2 f_{iK}(x - b_i)\, dx$$

The equality is trivial; the first inequality follows from the boundedness of $(\bar{m}_K'(x))^2$ (by Lemma 3(c) and the boundedness of $\left(m_\phi'(x) \right)^2$), the second inequality follows by $\Pr(A) \leq \Pr(B) + \Pr(C)$ whenever $A \subset (B \cap C)$, and the third inequality follows from Chebyshev's inequality and the variance bound of Theorem 1 (along with $\sigma_\varepsilon^2 > 0$). Of the terms in the final expression, we see immediately that the first term converges to zero by Assumption 7 and the second converges to zero by the uniform integrability of its integrand, which was shown in the course of the proof that Term II converges to zero, and by $d_K \to \infty$. The third term converges to zero by Lemmas 4 and 6(c).

Thus, Term Bi converges to zero.

Finally, turn to Term Bii. Consider the ith term of the average which makes up

Term Bii. We define $E_{D_i}^{out}\left[(\cdot) \mid b\right]$ to be $\int_{\hat{b}\in D_i\cap\left\{\hat{\sigma}_\varepsilon^2>2\sigma_\varepsilon^2\right\}}(\cdot)\, f_K\left(\hat{b} \mid b\right) d\hat{b}$ and $E_{D_i}^{cond}\left[(\cdot) \mid \hat{b}_i, \ b\right]$

to be $\int_{\hat{b}\in D_i}(\cdot)\, f_K\left(\hat{b}_{-i} \mid \hat{b}_i, b\right) d\hat{b}_{-i}$ (so that both the probability measure and the area

of integration depend on $\hat{b}_i$). Since $\left(\frac{a}{b}-\frac{c}{d}\right)^2 = \left(\frac{a}{b}\left(\frac{d-b}{d}\right)+\frac{a-c}{d}\right)^2 \le 2\left(\frac{a}{b}\right)^2\left(\frac{d-b}{d}\right)^2 +$

$2\left(\frac{a-c}{d}\right)^2$ We have:

$$E_{D_i}\left[\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2 \mid b\right] \tag{126a}$$

$$\le\quad 2E_{D_i}^{out}\left[\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right)^2 \mid b\right]+2E_{D_i}^{out}\left[\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2 \mid b\right]$$

$$+\int_{-d_K}^{d_K}E_{D_i}^{cond}\left[\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2 \mid \hat{b}_i, b\right] f_{iK}\left(\hat{b}_i-b_i\right) d\hat{b}_i$$

$$\le\quad 2C\left(q_K^2+s_K^{-2}\right)\Pr\left(\hat{\sigma}_\varepsilon^2>2\sigma_\varepsilon^2\right)$$

$$+2\int_{-d_K}^{d_K}\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2 E_{D_i}^{cond}\left[\left(\frac{\bar{m}_K\left(\hat{b}_i\right)-\hat{m}_{iK}\left(\hat{b}_i\right)}{\hat{m}_{iK}\left(\hat{b}_i\right)+s_K}\right)^2 \mid \hat{b}_i, b\right] f_{iK}\left(\hat{b}_i-b_i\right) d\hat{b}_i$$

$$+2\int_{-d_K}^{d_K}E_{D_i}^{cond}\left[\left(\frac{\hat{m}_{iK}'\left(\hat{b}_i\right)-\bar{m}_K'\left(\hat{b}_i\right)}{\hat{m}_{iK}\left(\hat{b}_i\right)+s_K}\right)^2 \mid \hat{b}_i, b\right] f_{iK}\left(\hat{b}_i-b_i\right) d\hat{b}_i$$

$$\le\quad 2\left(q_K^2+s_K^{-2}\right)\frac{C_2}{K}$$

$$+\frac{2}{s_K^2}\int_{-d_K}^{d_K}\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2 \times \tag{Term Biia}$$

$$E_{D_i}\left[\left(\bar{m}_K\left(\hat{b}_i\right)-\hat{m}_{iK}\left(\hat{b}_i\right)\right)^2 \mid \hat{b}_i, b\right] f_{iK}\left(\hat{b}_i-b_i\right) d\hat{b}_i$$

$$+\frac{2}{s_K^2}\int_{-d_K}^{d_K}E_{D_i}\left[\left(\hat{m}_{iK}'\left(\hat{b}_i\right)-\bar{m}_K'\left(\hat{b}_i\right)\right)^2 \mid \hat{b}_i, b\right] f_{iK}\left(\hat{b}_i-b_i\right) d\hat{b}_i. \tag{Term Biib}$$

where the first inequality follows from splitting the area of integration and $2a^2 + 2b^2 \geq (a+b)^2$ (and from the definition of d_K), the second inequality is due to the truncation of our score estimator and the boundedness of $(\bar{m}'_K(x))^2$ (by Lemma 3(c) and the boundedness of $\left(m'_\phi(x)\right)^2$), and applying the fact that $\left(\frac{a}{b} - \frac{c}{d}\right)^2 = \left(\frac{a}{b}\left(\frac{d-b}{d}\right) + \frac{a-c}{d}\right)^2 \leq 2\left(\frac{a}{b}\right)^2\left(\frac{d-b}{d}\right)^2 + 2\left(\frac{a-c}{d}\right)^2$ to the third term in the previous line. The final inequality is by Theorem 1 (for the first term) and is clear for the other two terms. The first term of the final line converges to zero (uniformly in i) due to Assumption 7, so we consider only the remaining two terms.

Term Biia satisfies

$$
\frac{2}{s_K^2} \int_{-d_K}^{d_K} \left(\frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right)^2 \times \tag{127}
$$
$$
E_{D_i}\left[\left(\bar{m}_K(\hat{b}_i) - \hat{m}_{iK}(\hat{b}_i)\right)^2 |\hat{b}_i, b\right] f_{iK}\left(\hat{b}_i - b_i\right) d\hat{b}_i
$$
$$
\leq \frac{2}{s_K^2}C\left(\frac{1}{h_K(K-1)} + h_K^2 + K^{-13/48}\log^2 K + 2K^{-13/96}h_K\log K + K^{-1/2}\log^2 K\right)
$$
$$
\times \int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right)^2 f_{iK}\left(\hat{b}_i - b_i\right) d\hat{b}_i
$$

by Lemma 6(a), since omitting the restriction to $\hat{b} \in D_i$ can only make the expectation larger, and conditioning on any b such that $\|b\|^2 \leq M < \infty$ will not change the results of that lemma, as can be easily verified. Also, Term Biib satisfies

$$
\frac{2}{s_K^2} \int_{-d_K}^{d_K} E_{D_i}\left[\left(\hat{m}'_{iK}(\hat{b}_i) - \bar{m}'_K(\hat{b}_i)\right)^2 |\hat{b}_i, \hat{b} \in D_i, b\right] f_{iK}\left(\hat{b}_i - b_i\right) d\hat{b}_i
$$
$$
\leq \frac{2}{s_K^2}C\left(\frac{1}{h_K^3(K-1)} + h_K^2 + K^{-5/48}\log^2 K + 2K^{-5/96}h_K\log K + K^{-1/4}\log^2 K\right)
$$

A-85

by Lemma 6(b), by the same logic. Thus we have that

$$
\text{Term Bii} \tag{128}
$$

$$
\leq \quad \frac{4\rho\sigma_\varepsilon^4}{s_K^2}C \left(
\begin{pmatrix} \frac{1}{h_K(K-1)} + h_K^2 + K^{-13/48}\log^2 K \\[4pt] + 2K^{-13/96}h_K\log K \; + \; K^{-1/2}\log^2 K \end{pmatrix} \\[6pt]
\times \int_{-\infty}^{\infty} \left(\frac{\bar{m}_K'(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right)^2 f_{iK}\left(\hat{b}_i - b_i\right) d\hat{b}_i \\[6pt]
+ \begin{pmatrix} \frac{1}{h_K^3(K-1)} + h_K^2 + K^{-5/48}\log^2 K \\[4pt] + 2K^{-5/96}h_K\log K \; + \; K^{-1/4}\log^2 K \end{pmatrix}
\right)
$$

$$
\rightarrow \quad 0
$$

by display (37) and Lemma 4 (after averaging), and we are finished with the proof

that $R\left(b,\ \hat{b}^{NSEB};\ f_K\right) - R\left(b,\ \hat{b}_{\tilde{G}_K}^{NB};\ \phi_K\right) \;\rightarrow\; 0$.

Now, to complete the proof that $\lim_{K\to\infty} \sup_{\|b\|_2 \leq M} \left| R\left(b,\ \hat{b}^{NSEB};\ \phi_K\right) - R\left(b,\ \hat{b}_{\tilde{G}_K}^{NB};\ \phi_K\right)\right| =$

0, we simply note that we never used any property of the vector b other than a bound

on the empirical variance of the b_i. Over the set $\{b : \|b\|_2 \leq M\}$ this bound is

uniform, so our results hold uniformly over this set. Thus, we simply substitute ϕ_K

in for f_K in the result derived above.

Further, we observe that

$$
\lim_{K\to\infty}\sup_{\|b\|_2 \leq M} \left\{ R\left(b,\ \hat{b}^{NSEB};\ \phi_K\right) - \inf_{\tilde{b}\in\mathcal{B}}\left\{R\left(b,\ \tilde{b};\ \phi_K\right)\right\}\right\} \tag{129}
$$

$$
= \quad \lim_{K\to\infty}\sup_{\|b\|_2 \leq M} \left\{
\begin{array}{l}
R\left(b,\ \hat{b}^{NSEB};\ \phi_K\right) - R\left(b,\ \hat{b}_{\tilde{G}_K}^{NB};\ \phi_K\right) + \\[6pt]
R\left(b,\ \hat{b}_{\tilde{G}_K}^{NB};\ \phi_K\right) - \inf_{\tilde{b}\in\mathcal{B}}\left\{R\left(b,\ \tilde{b};\ \phi_K\right)\right\}
\end{array}
\right\}
$$

A-86

$$\leq \quad \limsup_{K \to \infty} \sup_{\|b\|_2 \leq M} \left\{ R\left(b, \, \hat{b}^{NSEB}; \, \phi_K\right) - R\left(b, \, \hat{b}^{NB}_{\tilde{G}_K}; \, \phi_K\right) \right\}$$

$$+ \limsup_{K \to \infty} \sup_{\|b\|_2 \leq M} \left\{ R\left(b, \, \hat{b}^{NB}_{\tilde{G}_K}; \, \phi_K\right) - \inf_{\tilde{b} \in \mathcal{B}} \left\{ R\left(b, \, \tilde{b}; \, \phi_K\right) \right\} \right\}$$

$$\leq \quad \limsup_{K \to \infty} \sup_{\|b\|_2 \leq M} \left| R\left(b, \, \hat{b}^{NSEB}; \, \phi_K\right) - R\left(b, \, \hat{b}^{NB}_{\tilde{G}_K}; \, \phi_K\right) \right|$$

$$+ \limsup_{K \to \infty} \sup_{\|b\|_2 \leq M} \left\{ R\left(b, \, \hat{b}^{NB}_{\tilde{G}_K}; \, \phi_K\right) - \inf_{\tilde{b} \in \mathcal{B}} \left\{ R\left(b, \, \tilde{b}; \, \phi_K\right) \right\} \right\}$$

$$\leq \quad \limsup_{K \to \infty} \sup_{\|b\|_2 \leq M} \left\{ R\left(b, \, \hat{b}^{NB}_{\tilde{G}_K}; \, \phi_K\right) - \inf_{\tilde{b} \in \mathcal{B}} \left\{ R\left(b, \, \tilde{b}; \, \phi_K\right) \right\} \right\}$$

$$\leq \quad 0$$

where the second line follows from adding and subtracting $R\left(b, \, \hat{b}^{NB}_{\tilde{G}_K}; \, \phi_K\right)$, the third

line follows from the fact that $\limsup_{K \to \infty} \sup_x \{h_K(x) + g_K(x)\} \leq \limsup_{K \to \infty} \sup_x \{h_K(x)\} +$

$\limsup_{K \to \infty} \sup_x \{g_K(x)\}$, the fourth line follows from noting that $|x| \geq x$, the fifth

line follows from $\lim_{K \to \infty} \sup_{\|b\|_2 \leq M} \left| R\left(b, \, \hat{b}^{NSEB}; \, \phi_K\right) - R\left(b, \, \hat{b}^{NB}_{\tilde{G}_K}; \, \phi_K\right) \right| = 0$, and

the final line follows from $R\left(b, \, \hat{b}^{NB}_{\tilde{G}_K}; \, \phi_K\right) \leq \inf_{\tilde{b} \in \mathcal{B}} \left\{ R\left(b, \, \tilde{b}; \, \phi_K\right) \right\} \forall K$ and $\forall b$, as

shown in part (i) of this proof.

To bound the limsup of the sup of the absolute value, we must also consider the

liminf of the inf. To do so, observe that $\hat{b}^{NSEB}$ is an equivariant estimator, so that

$R\left(b, \, \hat{b}^{NSEB}; \, \phi_K\right) \geq \inf_{\tilde{b} \in \mathcal{B}} \left\{ R\left(b, \, \tilde{b}; \, \phi_K\right) \right\} \forall K$ and $\forall b$. Note that this domination

of $\hat{b}^{NSEB}$ may be by a different equivariant estimator at each b. This immediately

yields

$$- \liminf_{K \to \infty} \inf_{\|b\|_2 \leq M} \left\{ R\left(b, \, \hat{b}^{NSEB}; \, \phi_K\right) - \inf_{\tilde{b} \in \mathcal{B}} \left\{ R\left(b, \, \tilde{b}; \, \phi_K\right) \right\} \right\} \leq 0. \qquad (130)$$

which, together with our result above on the $\limsup_{K \to \infty} \sup_{\|b\|_2 \leq M}$, implies that

$$\limsup_{K \to \infty} \sup_{\|b\|_2 \leq M} \left| R\left(b, \hat{b}^{NSEB}; \phi_K\right) - \inf_{\tilde{b} \in \mathcal{B}} \left\{ R\left(b, \tilde{b}; \phi_K\right) \right\} \right| \tag{131}$$

$$\leq \quad \max \left\{ \begin{array}{l} -\liminf_{K \to \infty} \inf_{\|b\|_2 \leq M} \left\{ R\left(b, \hat{b}^{NSEB}; \phi_K\right) - \inf_{\tilde{b} \in \mathcal{B}} \left\{ R\left(b, \tilde{b}; \phi_K\right) \right\} \right\}, \\ \limsup_{K \to \infty} \sup_{\|b\|_2 \leq M} \left\{ R\left(b, \hat{b}^{NSEB}; \phi_K\right) - \inf_{\tilde{b} \in \mathcal{B}} \left\{ R\left(b, \tilde{b}; \phi_K\right) \right\} \right\} \end{array} \right\}$$

$$\leq \quad \max\{0, 0\}$$

$$\leq \quad 0$$

where the first inequality follows from the fact that $|x| = \max\{x, -x\}$ and the fact that we can interchange the maximum operation with the sup and limsup operations. The second inequality is by the bounds demonstrated above, and the final inequality is trivial. Now, since the limsup and sup operations are being performed on a nonnegative sequence, the result must be nonnegative. But by nonnegativity of the sequence again, the liminf of the sup must be nonnegative. Thus we obtain the desired conclusion:

$$\limsup_{K \to \infty} \sup_{\|b\|_2 \leq M} \left| R\left(b, \hat{b}^{NSEB}; \phi_K\right) - \inf_{\tilde{b} \in \mathcal{B}} \left\{ R\left(b, \tilde{b}; \phi_K\right) \right\} \right| = 0 \tag{132}$$

implies

$$\lim_{K \to \infty} \sup_{\|b\|_2 \leq M} \left| R\left(b, \hat{b}^{NSEB}; \phi_K\right) - \inf_{\tilde{b} \in \mathcal{B}} \left\{ R\left(b, \tilde{b}; \phi_K\right) \right\} \right| = 0.$$

(iii) The proof of part (iii) follows from:

$$\lim_{K \to \infty} \left\{ \sup_{\|b\|_2 \leq M} \left\{ \sup_{f_K} \left| R\left(b, \hat{b}^{NSEB}; f_K\right) - R\left(b, \hat{b}^{NB}_{\tilde{G}_K}; \phi_K\right) \right| \right\} \right\} = 0 \tag{133}$$

A-88

which is due to a procedure identical to that used in part (ii) Theorem 3, noting simply that the results there are uniform in the prior over a set of priors with the same variance bound. This uniformity is due to the fact that the only feature of the prior that is used is the variance of the prior or, in this frequentist case, the empirical variance of the b_i. Then we apply reasoning identical to that of the final portion of part (ii) above to conclude that

$$\lim_{K \to \infty} \sup_{\|b\|_2 \leq M} \sup_{f_K} \left| R\left(b, \hat{b}^{NSEB}; f_K\right) - \inf_{\tilde{b} \in \mathcal{B}} \left\{ R\left(b, \tilde{b}; \phi_K\right) \right\} \right| = 0 \tag{134}$$

as desired.

Q.E.D.

Proof of Theorem 6: As in the proof of Theorem 5, we note that, if we restrict $\|b\|_2 \leq M$ uniformly along the K sequence, we have that Lemmas 1 through 6 hold uniformly along the K sequence. But if Lemma 4 holds uniformly along the K sequence, then we get that, if $\tilde{G}_K \Longrightarrow G$, $\forall b$ s.t. $\|b\|_2 \leq M$ uniformly along the K sequence,

$$\lim_{K \to \infty} \left| R\left(b, \hat{b}^{NB}_{\tilde{G}_K}; \phi_K\right) - r_G\left(\hat{b}^{NB}, \phi\right) \right| = 0 \tag{135}$$

because

$$R\left(b, \hat{b}^{NB}_{\tilde{G}_K}; \phi_K\right) \tag{136}$$

$$= \rho \int_{-\infty}^{\infty} \left(\hat{b}^{NB}_{\tilde{G}_K,1}\left(\hat{b}_1\right) - b_1 \right)^2 d\tilde{G}_K\left(b_1\right)$$

A-89

$$= \rho \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\hat{b}_1 - \sigma_\varepsilon^2 \frac{m'_\phi\left(\hat{b}_1\right)}{m_\phi\left(\hat{b}_1\right)} - b_1 \right)^2 \phi\left(\hat{b}_1 - b_1\right) d\tilde{G}_K\left(b_1\right)$$

so if there is an uniformly integrable (with respect to $d\tilde{G}_K\left(b_1\right)$ $\forall K$ and $dG\left(b_1\right)$) dominating function of b_1 for $\int_{-\infty}^{\infty} \left(\hat{b}_1 - \sigma_\varepsilon^2 \frac{m'_\phi\left(\hat{b}_1\right)}{m_\phi\left(\hat{b}_1\right)} - b_1 \right)^2 \phi\left(\hat{b}_1 - b_1\right) d\hat{b}_1$, then we have the desired result by weak convergence (see, e. g., Billingsley (1968)). Now, $\left(\hat{b}_1 - \sigma_\varepsilon^2 \frac{m'_\phi\left(\hat{b}_1\right)}{m_\phi\left(\hat{b}_1\right)} - b_1 \right)^2$ $\leq$ $2\left(\hat{b}_1 - b_1\right)^2$ $+$ $2\sigma_\varepsilon^4 \left(\frac{m'_\phi\left(\hat{b}_1\right)}{m_\phi\left(\hat{b}_1\right)} \right)^2$, so that $\int_{-\infty}^{\infty} \left[2\left(\hat{b}_1 - b_1\right)^2 + 2\sigma_\varepsilon^4 \left(\frac{m'_\phi\left(\hat{b}_1\right)}{m_\phi\left(\hat{b}_1\right)} \right)^2 \right] \phi\left(\hat{b}_1 - b_1\right) d\hat{b}_1$ is a dominating function. This function is uniformly integrable w. r. t. $d\tilde{G}_K\left(b_1\right)$ $\forall K$ and $dG\left(b_1\right)$ by, for the first term, the fact that $\sigma_\varepsilon^2 < \infty$, and, for the second term, Lemma 4. Note that to apply Lemma 4, we must recognize that the second term doesn't depend on b_1, so that we may integrate with respect to $d\tilde{G}_K\left(b_1\right)$ or $dG\left(b_1\right)$ first, obtaining $\int_{-\infty}^{\infty} \frac{\left(m'_\phi\left(\hat{b}_1\right)\right)^2}{m_\phi\left(\hat{b}_1\right)} d\hat{b}_1$, which is uniformly bounded along the K sequence by Lemma 4 and the observation at the beginning of this proof, i. e., that Lemma 4 holds uniformly along the K sequence because $\|b\|_2 \leq M$ uniformly along the K sequence.

Q. E. D.

1   Appendix B: Berry-Esseen Theorems for Densities and Their Derivatives

This appendix provides Berry-Esseen-type theorems for densities and derivatives of densities for univariate and bivariate random variables. These theorems are referred to below as local limit results.

The presentation proceeds in two steps. First, the local limit results are proven assuming that a Berry-Esseen theorem (for c.d.f.'s) and a smoothness condition hold. Because Berry-Esseen theorems hold under a variety of primitive conditions, this provides general conditions under which the local limit results hold. Second, it is shown that a (multivariate) Berry-Esseen theorem does in fact hold for averages of strongly mixing random variables satisfying certain moment and mixing-rate conditions. This theorem is an adaptation of Tikhomirov's (1980) univariate result.

All theorems, lemmas, and equation numbers herein refer to this appendix only; this appendix is self-contained.

2   Local Limit Result

Let $\eta_1, \eta_2, \ldots$ be a sequence of m-dimensional random variables with mean zero and finite second moments. Without loss of generality, let $\lim_{n \to \infty} E\left[\left(\sqrt{\frac{1}{n}} \sum_{i=1}^{n} \eta_i\right)\left(\sqrt{\frac{1}{n}} \sum_{i=1}^{n} \eta_i\right)'\right] = I_m$. Let J_n denote the distribution function of $\sqrt{\frac{1}{n}} \sum_{i=1}^{n} \eta_i$ and let Φ denote the m-

variate standard normal distribution function. The local limit results will be proven under the following two conditions.

Condition A: The random variables $\eta_1, \eta_2, \ldots$ have conditional characteristic functions $\psi_1, \psi_2, \ldots$ (so that ψ_i is the characteristic function of the distribution of η_i conditional on η_j, $1 \leq j < i$) with the property that

$$\exists \, \alpha \; > \; 0, \; C_0 \; < \; \infty, \; M_0 < \infty \;\; s.t. \tag{1}$$

$$\sup_i |\psi_i(t)| \;\; \leq \;\; M_0 \, |t|^{-\alpha} \;\; \forall \, |t| \geq C_0.$$

Condition A is weaker than requiring, for instance, that the conditional densities of the η_i be uniformly bounded, or even that any of them be bounded, though it does rule out discreteness.

Condition B: A Berry-Esseen theorem holds for $\eta_1, \eta_2, \ldots$, that is,

$$\exists \, \beta > 0, \; \mu < \infty, \; M_1 < \infty \;\; s.t. \;\; \sup_{z \in R^m} |J_n(z) - \Phi(z)| \;\; \leq \;\; M_1 n^{-\beta} \log^\mu (n) \tag{2}$$

Typical Berry-Esseen theorems specify a particular constant M_1 and have $\beta = \frac{1}{2}$ and $\mu = 0$ (c. f. Feller (1971), F. Götze (1991), Hall and Heyde (1980)). However, the local limit results here do not rely on these specific values, and so are proven for the more general statement 2.

Lemma 1 (Cramér 1937) *Suppose $t, \zeta, \chi \in \Re^m$. If $\psi(t)$ is a characteristic function such that $|\psi(t)| \leq \nu < 1$ for all $|t| \geq M$, then we have for $|t| < M$*

$$|\psi(t)| \; \leq \; 1 - (1-\nu)^2 \, \frac{|t|^2}{8M^2} \tag{3}$$

Proof of Lemma 1: A terse proof for $m = 1$ is to be found on page 26 of Cramér (1937). However, it is given in an expanded form here for the reader's convenience.

Recall first that for scalars A and B, $\cos A \leq \frac{3}{4} + \frac{1}{4}\cos(2A)$ (since $\frac{1}{4}\cos(2A) - \cos(A) + \frac{3}{4} = \frac{1}{2}\cos^2(A) - \cos(A) + \frac{1}{2} = \frac{1}{2}(\cos(A) - 1)^2 \geq 0$) and $\sin(A - B) = \sin(A)\cos(B) - \sin(B)\cos(A)$.

$$
\begin{aligned}
|\psi(t)|^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{it'(\zeta - \chi)} dJ(\zeta)\, dJ(\chi) \qquad\qquad (4)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i\left[\sin(t'\zeta)\cos(t'\chi) - \sin(t'\chi)\cos(t'\zeta)\right] + \\
&\quad \left[\cos(t'(\zeta - \chi))\right] dJ(\zeta)\, dJ(\chi)\\
&= i\left[
\begin{array}{l}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sin(t'\zeta)\cos(t'\chi)\, dJ(\zeta)\, dJ(\chi) - \\
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sin(t'\chi)\cos(t'\zeta)\, dJ(\zeta)\, dJ(\chi)
\end{array}
\right] \\
&\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cos(t'(\zeta - \chi))\, dJ(\zeta)\, dJ(\chi)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cos(t'(\zeta - \chi))\, dJ(\zeta)\, dJ(\chi)\\
&\leq \frac{3}{4} + \frac{1}{4}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cos(2t'(\zeta - \chi))\, dJ(\zeta)\, dJ(\chi)\\
&= \frac{3}{4} + \frac{1}{4}\left\{
i\left[
\begin{array}{l}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sin(2t'\zeta)\cos(2t'\chi)\, dJ(\zeta)\, dJ(\chi) - \\
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sin(2t'\chi)\cos(2t'\zeta)\, dJ(\zeta)\, dJ(\chi)
\end{array}
\right] +
\right.\\
&\quad \left.
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cos(2t'(\zeta - \chi))\, dJ'(\zeta)\, dJ(\chi)
\right\}\\
&= \frac{3}{4} + \frac{1}{4}|\psi(2t)|^2
\end{aligned}
$$

where the first equality holds because the square of a complex number's modulus equals its product (in complex multiplication) with its complex conjugate, the sec-

B-3

ond equality holds by the trigonometric identity recalled above, the third is merely

a rearrangement, and the fourth follows from the fact that the two integrals in the

imaginary part are equal, so their difference is zero (we may simply relabel the vari-

ables). The inequality comes from the trigonometric inequality noted above, and

the last two equalities are simply the first four "in reverse." Thus, we have that for

$|t| \in \left[\frac{M}{2}, M \right)$

$$|\psi(t)|^2 \leq 1 - \frac{1}{4}(1 - \nu)^2 \tag{5}$$

and we may repeat this argument to show that for $|t| \in \left[\frac{M}{2^q}, \frac{M}{2^{q-1}} \right)$ (for any integer

$q \geq 1$)

$$|\psi(t)|^2 \leq 1 - \left(\frac{1}{4}\right)^q (1 - \nu)^2 < 1 - (1 - \nu)^2 \frac{|t|^2}{4M^2} \tag{6}$$

so that, in the same region,

$$|\psi(t)| \leq 1 - (1 - \nu)^2 \frac{|t|^2}{8M^2} \tag{7}$$

as can be seen by squaring the righthand side of the inequality immediately above

and comparing it to the rightmost quantity of the inequality in the previous display.

Now, q is arbitrary, so the desired conclusion must hold for $|t| \in (0, M)$. But we

know that $\psi(0) = 1$, so we are finished. Q.E.D.

Theorem 2 (Univariate Local Limit Theorem) *Suppose that Conditions A and*

B hold, and let $m = 1$. Then $\forall d \in N$, $\exists B(d) < \infty$ and $n_0(d) < \infty$ such that,

$\forall n > n_0\left(d\right),$

$$\sup_{z \in R} \left| j_n^{(d)}\left(z\right) - \phi^{(d)}\left(z\right) \right| \ \leq \ B\left(d\right) n^{-\beta \,/\, 2^{d+1}} \left[\log\left(n\right)\right]^{\mu \,/\, 2^{d+1}}$$

where $j_n^{(d)}$ is the dth derivative, $d = 0, 1, 2, \ldots$, of the density of $\frac{\eta_1 + \eta_2 + \cdots + \eta_n}{\sqrt{n}}$ (this density will be shown to exist for sufficiently large n as part of the proof), and $\phi^{(d)}$ is the dth derivative of the standard normal density.

Proof of Theorem 2: We will proceed by induction on the order d of the derivative to be taken. First we will prove that the result holds for $d = 0$. Note that if the first derivative of j_n is uniformly bounded for all n greater than some given n_0, then j_n will clearly satisfy a Lipschitz condition uniformly beyond n_0, that is, we will have $\sup_{z,w \in \Re} \sup_{n \geq n_0} \left| j_n\left(z\right) - j_n\left(w\right) \right| \ \leq \ B_0 \left| z - w \right|$ for some $B_0 < \infty$. Clearly, ϕ satisfies such a smoothness condition (all its derivatives exist and are bounded). Let $B_1 < \infty$ be the Lipschitz constant for ϕ. Now, if the Lipschitz constant holds for j_n,

$$
\begin{aligned}
sup_{z \in R} \left| j_n\left(z\right) - \phi\left(z\right) \right| \ \leq \ & \sup_{z \in R} \left| j_n\left(z\right) - \frac{\int_z^{z+r} j_n\left(w\right) dw}{r} \right| \ + \qquad\qquad (8)\\
& \sup_{z \in R} \frac{1}{r} \left| \int_z^{z+r} j_n\left(w\right) dw - \int_z^{z+r} \phi\left(w\right) dw \right| \ + \\
& \sup_{z \in R} \left| \phi\left(z\right) - \frac{\int_z^{z+r} \phi\left(w\right) dw}{r} \right| \\
\leq \ & \sup_{z \in R} \left| j_n\left(z\right) - j_n\left(c_f\right) \right| \ + \\
& \sup_{z \in R} \frac{1}{r} \left| \int_z^{z+r} j_n\left(w\right) dw - \int_z^{z+r} \phi\left(w\right) dw \right| \ + \\
& \sup_{z \in R} \left| \phi\left(z\right) - \phi\left(c_\phi\right) \right|
\end{aligned}
$$

B-5

$$\leq \quad B_0 \left| z - c_f \right| +$$

$$\frac{1}{r} \sup_{z \in \Re} \left| J_n(z+r) - J_n(z) - \Phi(z+r) + \Phi(z) \right| +$$

$$B_1 \left| z - c_\phi \right|$$

$$\leq \quad 2 \left(B_0 + B_1 \right) r + \frac{2}{r} \sup_{z \in \Re} \left| J_n(z) - \Phi(z) \right|$$

$$\leq \quad \left(B_0 + B_1 \right) r + \frac{2M_1}{rn^\beta} \log^\mu(n)$$

$$\leq \quad 2 \left[2^{1/2} \left(B_0 + B_1 \right)^{1/2} M_1^{1/2} \right] n^{-\beta/2} \log^{\mu/2}(n)$$

where $c_f, c_\phi \in [z, z+r]$ by the mean value theorem. The second term in the fifth in-equality uses Condition B, and the final inequality follows by setting $r = \sqrt{\frac{2M_1}{B_0+B_1}} n^{-\beta/2} \log^{\mu/2}(n)$, which is certainly permitted, since r is arbitrary.

Thus the lemma is proven with $B(0) = 2 \left[2^{1/2} \left(B_0 + B_1 \right)^{1/2} M_1^{1/2} \right]$ if we can show that the Lipschitz condition holds.

This is where Cramér's (1937) lemma is useful: consider general $d \in N$. By the Fourier inversion theorem,

$$
\begin{aligned}
\left| j_n^{(d)}(z) \right| \quad &\leq \quad \int_{-\infty}^{\infty} |t|^d \left\{ \prod_{i=1}^{n} \left| \psi_i \left(\frac{t}{\sqrt{n}} \right) \right| \right\} dt \qquad (9)\\
&\leq \quad \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |t|^d \left\{ \prod_{i=1}^{n} \left| \psi_i \left(\frac{t}{\sqrt{n}} \right) \right| \right\} dt + \int_{\delta\sqrt{n}}^{\infty} |t|^d \left\{ \prod_{i=1}^{n} \left| \psi_i \left(\frac{t}{\sqrt{n}} \right) \right| \right\} dt\\
&\quad + \int_{-\infty}^{-\delta\sqrt{n}} |t|^d \left\{ \prod_{i=1}^{n} \left| \psi_i \left(\frac{t}{\sqrt{n}} \right) \right| \right\} dt.
\end{aligned}
$$

Now choose $\delta = \max \left\{ (2M_0)^{1/\alpha}, C_0 \right\}$, where the parameters refer to Condition A. Then, by Condition A, $\forall t > \delta$, $\sup_i |\psi_i(t)| \leq \frac{1}{2} \equiv \nu$. We can now apply Cramér's

lemma to obtain $\sup_i |\psi_i(t)| \leq 1 - (1-\nu)^2 \frac{|t|^2}{8\delta^2} \ \forall \ |t| < \delta$. Thus, we have,

$$
\sup_{n \geq n_0} \left| j_n^{(d)}(z) \right| \ \leq \ \sup_{n \geq n_0} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |t|^d \left(1 - (1-\nu)^2 \frac{t^2 / n}{8\delta^2} \right)^n dt \ + \tag{10}
$$

$$
\sup_{n \geq n_0} \nu^{n-m} \int_{\delta\sqrt{n}}^{\infty} |t|^d \left(\frac{M_0}{|t / \sqrt{n}|^\alpha} \right)^m dt \ +
$$

$$
\sup_{n \geq n_0} \nu^{n-m} \int_{-\infty}^{-\delta\sqrt{n}} |t|^d \left(\frac{M_0}{|t / \sqrt{n}|^\alpha} \right)^m dt
$$

$$
\leq \ \sup_{n \geq n_0} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |t|^d \exp\left(-Ct^2\right) dt \ +
$$

$$
2 \sup_{n \geq n_0} \nu^{n-m} M_0^m n^{\alpha m / 2} \int_{-\infty}^{-\delta\sqrt{n}} |t|^{d-\alpha m} dt
$$

$$
\leq \ \sup_{n \geq n_0} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |t| \exp\left(-Ct^2\right) dt \ +
$$

$$
2 \sup_{n \geq n_0} M_0^{\lceil (d+2) / \alpha \rceil} \nu^{n - \lceil (d+2) / \alpha \rceil} n^{(d+2) / 2} \int_{\delta}^{\infty} |t|^{-2} dt
$$

$$
\leq \ M_2(d) \ < \ \infty
$$

where the second inequality follows by setting $C = \frac{(1-\nu)^2}{8\delta^2}$ and observing that $\exp\left(-Ct^2\right) = \sup_n \left(1 - \frac{Ct^2}{n}\right)^n$, the third inequality follows by choosing $m = \left\lceil \frac{d+2}{\alpha} \right\rceil$, where $\lceil \cdot \rceil$ is the least greater (or equal) integer function. The desired Lipschitz condition is now verified. We have only to set $d = 0$ to prove the existence and boundedness of the density itself for sufficiently large n in an identical fashion.

We have now proven the $d = 0$ case. To prove higher-d cases, we now simply substitute $j_n^{(d)}$ for j_n and $\phi^{(d)}$ for ϕ in (8) and, following the steps in (8), we obtain

$$
\sup_{z \in \Re} \left| j_n^{(d)}(z) - \phi^{(d)}(z) \right| \tag{11}
$$

$$
\leq \ (B_0(d) + B_1(d)) r \ + \ \frac{2}{r} \sup_{z \in \Re} \left| j_n^{(d-1)}(z) - \phi^{(d-1)}(z) \right|
$$

where the Lipschitz condition holds as a consequence of (10). It is readily verified that the bound in the statement of the theorem satisfies the recursion (11), and since we have shown the bound for $d = 0$, we are finished. Q. E. D.

Theorem 3 (Bivariate Local Limit Theorem) *Suppose that Conditions A and B hold with $m = 2$. Let $j_n^{(0)}(z, w)$ denote the density of $\sqrt{\frac{1}{n}} \sum_{i=1}^{n} \eta_i$, and let $\phi(z, w)$ denote the bivariate standard normal density. Then $\forall\, d \in N,\ \exists\, B(d) < \infty$ such that*

$$\sup_{z \in R, w \in R} \left| j_n^{(d)}(z, w) - \phi^{(d)}(z, w) \right| \leq B(d)\, n^{-\beta\,/\,(3 \times 2^d)} \left[\log(n)\right]^{\mu\,/\,(3 \times 2^d)}$$

where $j_n^{(d)}$ is the dth derivative, $d = 0, 1, 2, \ldots$, of $j_n^{(0)}(z, w)$ with respect to z (this density will be shown to exist for sufficiently large n as part of the proof) and $\phi^{(d)}$ is the dth derivative of $\phi(z, w)$ with respect to z.

Proof of Theorem 3: We will proceed by induction on the order d of the derivative to be taken. First we will prove that the result holds for $d = 0$. Note that if the first derivative of j_n with respect to z is uniformly bounded for all n greater than some given n_0, then j_n will clearly satisfy a Lipschitz condition in z uniformly beyond n_0, that is, we will have $\sup_{z,w,u,v \in \Re} \sup_{n \geq n_0} |j_n(z, w) - j_n(u, v)| \leq B_0 \left\| \begin{pmatrix} z \\ w \end{pmatrix} - \begin{pmatrix} u \\ v \end{pmatrix} \right\|$ and for some $B_0 < \infty$. Now, if such a uniform Lipschitz condition holds, we can easily show that, since ϕ clearly satisfies such a smoothness condition (all its derivatives

exist and are bounded),

$$sup_{z,w\in R}\left|j_n\left(z,w\right)-\phi\left(z,w\right)\right| \tag{12}$$

$$\leq \quad \sup_{z,w\in\Re}\left|j_n\left(z,w\right)-\frac{\int_z^{z+r}\int_w^{w+r}j_n\left(u,v\right)dudv}{r^2}\right| \;+$$

$$\sup_{z,w\in\Re}\frac{1}{r^2}\left|\int_z^{z+r}\int_w^{w+r}j_n\left(u,v\right)dudv-\int_z^{z+r}\int_w^{w+r}\phi\left(u,v\right)dudv\right| \;+$$

$$\sup_{z,w\in\Re}\left|\phi\left(z,w\right)-\frac{\int_z^{z+r}\int_w^{w+r}\phi\left(u,v\right)dudv}{r^2}\right|$$

$$\leq \quad \sup_{z,w\in\Re}\left|j_n\left(z,w\right)-j_n\left(c_f,d_f\right)\right| \;+$$

$$\sup_{z,w\in\Re}\frac{1}{r^2}\left|\int_z^{z+r}\int_w^{w+r}j_n\left(u,v\right)dudv-\int_z^{z+r}\int_w^{w+r}\phi\left(u,v\right)dudv\right| \;+$$

$$\sup_{z,w\in\Re}\left|\phi\left(z,w\right)-\phi\left(c_\phi,d_\phi\right)\right|$$

$$\leq \quad B_0\left\|\begin{pmatrix}z\\w\end{pmatrix}-\begin{pmatrix}c_f\\d_f\end{pmatrix}\right\| + \frac{2}{r^2}\sup_{z,w\in\Re}\left|J_n\left(z,w\right)-\Phi\left(z,w\right)\right| \;+$$

$$B_1\left\|\begin{pmatrix}z\\w\end{pmatrix}-\begin{pmatrix}c_\phi\\d_\phi\end{pmatrix}\right\|$$

$$\leq \quad \left(B_0+B_1\right)r \;+\; \frac{2M_1}{r^2n^\beta}\log^\mu\left(n\right)$$

$$\leq \quad \left[\left(4^{1/3}+4^{1/6}\right)\left(B_0+B_1\right)^{2/3}M_1^{1/3}\right]n^{-\beta/3}\log^{\mu/3}\left(n\right)$$

where $c_f,c_\phi\in\left[z,z+r\right]$ and $d_f,d_\phi\in\left[w,w+r\right]$ by the mean value theorem. The

second term in the fourth inequality uses Condition B, and the final inequality follows

by setting $r=\sqrt[3]{\frac{4M_1}{B_0+B_1}}n^{-\beta/3}\log^{\mu/3}\left(n\right)$, which is certainly permitted, since r is

arbitrary.

Thus the lemma is proven with $B(0) = \left[\left(4^{1/3} + 4^{1/6}\right)(B_0 + B_1)^{2/3} M_1^{1/3}\right]$ if we can show that the Lipschitz condition holds.

This is where Cramér's (1937) lemma is useful: consider general $d \in N$. By the Fourier inversion theorem,

$$\left|j_n^{(d)}(z,w)\right| \;\leq\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |t_1|^d \left\{\prod_{i=1}^{n}\left|\psi_i\left(\frac{t}{\sqrt{n}}\right)\right|\right\} dt_1 dt_2 \tag{13}$$

$$\leq\; \int_{B_{\delta\sqrt{n}}} |t_1|^d \left\{\prod_{i=1}^{n}\left|\psi_i\left(\frac{t}{\sqrt{n}}\right)\right|\right\} dt \;+\; \int_{\Re^2 - B_{\delta\sqrt{n}}} |t|^d \left\{\prod_{i=1}^{n}\left|\psi_i\left(\frac{t}{\sqrt{n}}\right)\right|\right\} dt$$

where $B_{\delta\sqrt{n}} \equiv \{t \;:\; |t| \leq \delta\sqrt{n}\}$, and, by choosing $\delta = \max\left\{(2M_0)^{1/\alpha}, C_0\right\}$, we will have $\nu = \frac{1}{2}$ such that $\sup_i |\psi_i(t)| \leq \nu \; \forall \; |t| \geq \delta$. But then Cramér's lemma proves that we have $\sup_i |\psi_i(t)| \leq 1 - (1-\nu)^2 \frac{|t|^2}{8\delta^2} \; \forall \; |t| < \delta$. Thus, we have

$$\sup_{n \geq n_0} \left|j_n^{(d)}(z,w)\right| \tag{14}$$

$$\leq\; \sup_{n \geq n_0} \int_{B_{\delta\sqrt{n}}} |t_1|^d \left(1 - (1-\nu)^2 \frac{|t|^2/n}{8\delta^2}\right)^n dt \;+$$

$$\sup_{n \geq n_0} \nu^{n-m} \int_{\Re^2 - B_{\delta\sqrt{n}}} |t_1|^d \left(\frac{M_0}{|t/\sqrt{n}|^\alpha}\right)^m dt$$

$$\leq\; \sup_{n \geq n_0} \int_{B_{\delta\sqrt{n}}} |t_1|^d \exp\left(-C|t|^2\right) dt \;+$$

$$\sup_{n \geq n_0} M_0^m \nu^{n-m} n^{\alpha m/2} \int_{\Re^2 - B_{\delta\sqrt{n}}} |t_1|^d |t|^{-\alpha m} dt$$

$$\leq\; \sup_{n \geq n_0} \int_{B_{\delta\sqrt{n}}} |t_1|^d \exp\left(-C|t|^2\right) dt \;+$$

$$\sup_{n \geq n_0} \dot{M}_0^{\lceil (d+4)/\alpha \rceil} \nu^{n - \lceil (d+4)/\alpha \rceil} n^{(d+4)/2} \int_{\Re^2 - B_{\delta\sqrt{n}}} |t_1|^d |t|^{-(d+4)} dt$$

$$\leq\; \sup_{n \geq n_0} \int_{\Re^2} |t_1|^d \exp\left(-C|t|^2\right) dt \;+$$

$$\sup_{n \geq n_0} M_0^{\lceil (d+4)/\alpha \rceil} \nu^{n - \lceil (d+4)/\alpha \rceil} n^{(d+4)/2} \int_{\Re^2 - B_{\delta\sqrt{n}}} |t|^{-4} dt$$

$$\leq \quad \sup_{n \geq n_0} \int_{\Re^2} |t_1| \exp\left(-C\,|t|^2\right) dt \quad +$$

$$\sup_{n \geq n_0} M_0^{\lceil (d+4)\,/\,\alpha \rceil} \nu^{n-\lceil (d+4)\,/\,\alpha \rceil} n^{(d+4)\,/\,2} \int_{\Re^2 - B_{\delta\sqrt{n}}} r^{-3} dr d\theta$$

$$\leq \quad M_2 \; + \; \sup_{n \geq n_0} M_0^{\lceil (d+4)\,/\,\alpha \rceil} \nu^{n-\lceil (d+4)\,/\,\alpha \rceil} n^{(d+2)\,/\,2} 2\pi \int_{\delta}^{\infty} r^{-3} dr$$

$$\leq \quad M_3 \; < \; \infty$$

where the second inequality follows by setting $C = \frac{(1-\nu)^2}{8\delta^2}$ and noting that $\exp\left(-C\,|t|^2\right) = \sup_n \left(1 - \frac{(1-\nu)^2 |t|^2}{8\delta^2 n}\right)^n$, and the third inequality follows by setting $m = \left\lceil \frac{d+4}{\alpha} \right\rceil$, where

$\lceil \cdot \rceil$ is the least integer greater function. The fourth inequality is due to the fact that

$|t_1|^2 = \left((1\ 0\ \cdots\ 0)\,t\right)^2 \leq |t|^2$ by the Cauchy-Schwartz inequality, so that $|t_1| \leq |t|$,

and the fifth inequality follows upon a change of variables into polar coordinates. So

the dth derivative of j_n w. r. t. z is bounded. An identical proof in which $|t_2|$ is

substituted for $|t_1|$ shows that the dth derivative of j_n w. r. t. w is bounded. To-

gether, these two bounds produce the Lipschitz condition. We have only to set $d = 0$

in the proof above to prove the existence and boundedness of the density itself for

sufficiently large n.

We have now proven the $d = 0$ case. To prove higher-d cases, we now simply

substitute $j_n^{(d)}$ for j_n and $\phi^{(d)}$ for ϕ in (12), and, only slightly modifying the steps in

(12), we obtain

$$sup_{z,w \in R} \left| j_n^{(d)}(z,w) - \phi^{(d)}(z,w) \right| \tag{15}$$

$$\leq \quad \sup_{z,w \in \Re} \left| j_n^{(d)}(z,w) - \frac{\int_z^{z+r} j_n^{(d)}(u,w)\,du}{r} \right| +$$

$$\sup_{z,w\in\Re}\frac{1}{r}\left|\int_z^{z+r}j_n^{(d)}\left(u,w\right)du-\int_z^{z+r}\phi^{(d)}\left(u,w\right)du\right|+$$

$$\sup_{z,w\in\Re}\left|\phi\left(z,w\right)-\frac{\int_z^{z+r}\phi^{(d)}\left(u,w\right)du}{r}\right|$$

$$\leq\quad\left(B_0\left(d\right)+B_1\left(d\right)\right)r\ +\ \frac{2}{r}\sup_{z,w\in\Re}\left|j_n^{(d-1)}\left(z,w\right)-\phi^{(d-1)}\left(z,w\right)\right|.$$

Note that the second term of the upper bound is, as a function of r, only an inverse, rather than an inverse squared, because we have taken only a first partial derivative, rather than a cross or second partial. It is readily verified that the bound in the statement of the theorem satisfies the recursion (15), and since we have shown the bound for $d=0$, we are finished. Q. E. D.

3   Multivariate Berry-Esseen Theorem Under Strong Mixing

We now provide a multivariate Berry-Esseen theorem that applies to sequences $\{\eta_i\}_{i=1}^{\infty}$ of random variables in $\Re^m$ which satisfy a strong mixing condition and a moment condition.

Definition 1 *A sequence of random variables $\eta_1, \eta_2, \ldots$ will be said to be strongly mixing with coefficients $\alpha\left(n\right)$ if*

$$\alpha\left(n\right)\ =\ \sup_{k,A\in\mathcal{F}_1^k, B\in\mathcal{F}_{k+n}^{\infty}}\left|P\left(AB\right)-P\left(A\right)P\left(B\right)\right| \tag{16}$$

where $\mathcal{F}_a^b$ is the σ-algebra generated by η_j, $j\in\{a, a+1, \ldots b\}$.

Condition C. Let $\eta_1, \eta_2, \ldots$ be a sequence of $\Re^m$-valued random variables with $E[\eta_i] = 0$, $\sup_i E\left[|\eta_i|^{4+\gamma}\right] < \infty$ for some $\gamma > 0$, and $\alpha(n) \leq M_3 e^{-\beta n}$ for some $M_3 < \infty$ and some $\beta > 0$.

Theorem 4 provides a m-variate Berry-Esseen theorem which holds under Condition C. The results of this theorem satisfy Condition B. Thus the local limit results given above hold under, in particular, Conditions A and C.

This section concludes with the statement of this theorem, which is minor modification of a result of Tikhomirov (1980).

Theorem 4 *Suppose Condition C is satisfied. Then there is a constant C_2 depending only on m, β, γ, M_3 such that*

$$\sup_{z \in \Re^m} |J_n(z) - \Phi(z)| \leq C_2 n^{-1/2} \log n \tag{17}$$

where $J_n(z) = \Pr(S_{n,1} \leq z_1, S_{n,2} \leq z_2, \ldots, S_{n,m} \leq z_m)$, *with* $S_n = H_n^{-1} \sum_{i=1}^n \eta_i$, *where* $H_n \equiv Cholesky\left\{E\left[\left(\sum_{i=1}^n \eta_i\right)\left(\sum_{i=1}^n \eta_i\right)'\right]\right\}$ *(the Cholesky factor of the given expectation).*

Proof: Consider Tikhomirov's (1980) Theorem 4. Although this is a univariate result, we see that minor modifications make it applicable to $\Re^m$. Namely, the ODE that Tikhomirov derives for the characteristic function becomes a system of ODEs, and the solution, naturally, becomes a multivariate characteristic function. However,

the structure of his proof remains exactly the same; his lemmas transfer naturally to the multivariate case.

Appendix C: Description of Time Series Data Used in the Empirical Analysis

This appendix lists the time series used to construct the forecasts discussed in section 4. The format is: series number; series mnemonic; data span used; transformation code; and brief series description. The transformation codes are: 1 = no transformation; 2 = first difference; 4 = logarithm; 5 = first difference of logarithms; 6 = second difference of logarithms. An asterisk after the date denotes a series that was included in the unbalanced panel but not the balanced panel, either because of missing data or because of gross outliers which were treated as missing data. The series were either taken directly from the DRI-McGraw Hill Basic Economics database, in which case the original mnemonics are used, or they were produced by authors' calculations based on data from that database, in which case the authors calculations and original DRI/McGraw series mnemonics are provided. The following abbreviations appear in the data definitions: SA = seasonally adjusted; NSA = not seasonally adjusted; SAAR = seasonally adjusted at an annual rate; FRB = Federal Reserve Board; AC = Authors calculations

Real output and income

ip	1959:01-1998:12	5	industrial production: total index (1992=100,sa)
ipp	1959:01-1998:12	5	industrial production: products, total (1992=100,sa)
ipf	1959:01-1998:12	5	industrial production: final products (1992=100,sa)
ipc	1959:01-1998:12	5	industrial production: consumer goods (1992=100,sa)
ipcd	1959:01-1998:12	5	industrial production: durable consumer goods (1992=100,sa)
ipcn	1959:01-1998:12	5	industrial production: nondurable condsumer goods (1992=100,sa)
ipe	1959:01-1998:12	5	industrial production: business equipment (1992=100,sa)
ipi	1959:01-1998:12	5	industrial production: intermediate products (1992=100,sa)
ipm	1959:01-1998:12	5	industrial production: materials (1992=100,sa)
ipmd	1959:01-1998:12	5	industrial production: durable goods materials (1992=100,sa)
ipmnd	1959:01-1998:12	5	industrial production: nondurable goods materials (1992=100,sa)
ipmfg	1959:01-1998:12	5	industrial production: manufacturing (1992=100,sa)
ipd	1959:01-1998:12	5	industrial production: durable manufacturing (1992=100,sa)
ipn	1959:01-1998:12	5	industrial production: nondurable manufacturing (1992=100,sa)
ipmin	1959:01-1998:12	5	industrial production: mining (1992=100,sa)
iput	1959:01-1998:12	5	industrial production: utilities (1992-=100,sa)
ipxmca	1959:01-1998:12	1	capacity util rate: manufacturing,total(% of capacity,sa)(frb)
pmi	1959:01-1998:12	1	purchasing managers' index (sa)
pmp	1959:01-1998:12	1	NAPM production index (percent)
gmyxpq	1959:01-1998:12	5	personal income less transfer payments (chained) (#51) (bil 92$,saar)

Employment and hours

lhel	1959:01-1998:12	5	index of help-wanted advertising in newspapers (1967=100;sa)
lhelx	1959:01-1998:12	4	employment: ratio; help-wanted ads:no. unemployed clf
lhem	1959:01-1998:12	5	civilian labor force: employed, total (thous.,sa)
lhnag	1959:01-1998:12	5	civilian labor force: employed, nonagric.industries (thous.,sa)
lhur	1959:01-1998:12	1	unemployment rate: all workers, 16 years & over (%,sa)
lhu680	1959:01-1998:12	1	unemploy.by duration: average(mean)duration in weeks (sa)
lhu5	1959:01-1998:12	1	unemploy.by duration: persons unempl.less than 5 wks (thous.,sa)
lhu14	1959:01-1998:12	1	unemploy.by duration: persons unempl.5 to 14 wks (thous.,sa)
lhu15	1959:01-1998:12	1	unemploy.by duration: persons unempl.15 wks + (thous.,sa)
lhu26	1959:01-1998:12	1	unemploy.by duration: persons unempl.15 to 26 wks (thous.,sa)
lpnag	1959:01-1998:12	5	employees on nonag. payrolls: total (thous.,sa)
lp	1959:01-1998:12	5	employees on nonag payrolls: total, private (thous,sa)
lpgd	1959:01-1998:12	5	employees on nonag. payrolls: goods-producing (thous.,sa)
lpcc	1959:01-1998:12	5	employees on nonag. payrolls: contract construction (thous.,sa)

lpem	1959:01-1998:12 5	employees on nonag. payrolls: manufacturing (thous.,sa)
lped	1959:01-1998:12 5	employees on nonag. payrolls: durable goods (thous.,sa)
lpen	1959:01-1998:12 5	employees on nonag. payrolls: nondurable goods (thous.,sa)
lpsp	1959:01-1998:12 5	employees on nonag. payrolls: service-producing (thous.,sa)
lpt	1959:01-1998:12 5	employees on nonag. payrolls: wholesale & retail trade (thous.,sa)
lpfr	1959:01-1998:12 5	employees on nonag. payrolls: finance,insur.&real estate (thous.,sa
lps	1959:01-1998:12 5	employees on nonag. payrolls: services (thous.,sa)
lpgov	1959:01-1998:12 5	employees on nonag. payrolls: government (thous.,sa)
lphrm	1959:01-1998:12 1	avg. weekly hrs. of production wkrs.: manufacturing (sa)
lpmosa	1959:01-1998:12 1	avg. weekly hrs. of prod. wkrs.: mfg.,overtime hrs. (sa)
pmemp	1959:01-1998:12 1	NAPM employment index (percent)

Real retail, manufacturing and trade sales

msmtq	1959:01-1998:12 5	manufact. & trade: total (mil of chained 1992 dollars)(sa)
msmq	1959:01-1998:12 5	manufact. & trade:manufacturing;total(mil of chained 1992 dollars)(sa)
msdq	1959:01-1998:12 5	manufact. & trade:mfg; durable goods (mil of chained 1992 dollars)(sa)
msnq	1959:01-1998:12 5	manufact. & trade:mfg;nondurable goods (mil of chained 1992 dollars)(sa)
wtq	1959:01-1998:12 5	merchant wholesalers: total (mil of chained 1992 dollars)(sa)
wtdq	1959:01-1998:12 5	merchant wholesalers:drble goods total (mil of chained 1992 dollars)(sa)
wtnq	1959:01-1998:12 5	merchant wholesalers:nondurable goods (mil of chained 1992 dollars)(sa)
rtq	1959:01-1998:12 5	retail trade: total (mil of chained 1992 dollars)(sa)
rtnq	1959:01-1998:12 5	retail trade:nondurable goods (mil of 1992 dollars)(sa)

Consumption

gmcq	1959:01-1998:12 5	personal consumption expend (chained)-total (bil 92$,saar)
gmcdq	1959:01-1998:12 5	personal consumption expend (chained)-total durables (bil 92$,saar)
gmcnq	1959:01-1998:12 5	personal consumption expend (chained)-nondurables (bil 92$,saar)
gmcsq	1959:01-1998:12 5	personal consumption expend (chained)-services (bil 92$,saar)
gmcanq	1959:01-1998:12 5	personal cons expend (chained)-new cars (bil 92$,saar)

Housing starts and sales

hsfr	1959:01-1998:12 4	housing starts:nonfarm(1947-58);total farm&nonfarm(1959-)(thous.,sa
hsne	1959:01-1998:12 4	housing starts:northeast (thous.u.)s.a.
hsmw	1959:01-1998:12 4	housing starts:midwest(thous.u.)s.a.
hssou	1959:01-1998:12 4	housing starts:south (thous.u.)s.a.
hswst	1959:01-1998:12 4	housing starts:west (thous.u.)s.a.
hsbr	1959:01-1998:12 4	housing authorized: total new priv housing units (thous.,saar)
hmob	1959:01-1998:12 4	mobile homes: manufacturers' shipments (thous.of units,saar)
hsbne	1959:01-1998:12 4	houses authorized by build. permits:northeast(thou.u.)s.a
hsbmw	1959:01-1998:12 4	houses authorized by build. permits:midwest(thou.u.)s.a
hsbsou	1959:01-1998:12 4	houses authorized by build. permits:south(thou.u.)s.a
hsbwst	1959:01-1998:12 4	houses authorized by build. permits:west(thou.u.)s.a

Real inventories and inventory-sales ratios

ivmtq	1959:01-1998:12 5	manufacturing & trade inventories:total (mil of chained 1992)(sa)
ivmfgq	1959:01-1998:12 5	inventories, business, mfg (mil of chained 1992 dollars, sa)
ivmfdq	1959:01-1998:12 5	inventories, business durables (mil of chained 1992 dollars, sa)
ivmfnq	1959:01-1998:12 5	inventories, business, nondurables (mil of chained 1992 dollars, sa)
ivwrq	1959:01-1998:12 5	manufact. & trade inv: merchant wholesalers (mil of chained 1992 dollars)(sa)
ivrrq	1959:01-1998:12 5	manufacturing & trade inv:retail trade (mil of chained 1992 dollars)(sa)
ivsrq	1959:01-1998:12 2	ratio for mfg & trade: inventory/sales (chained 1992 dollars, sa)
ivsrmq	1959:01-1998:12 2	ratio for mfg & trade:mfg;inventory/sales (87$)(s.a.)
ivsrwq	1959:01-1998:12 2	ratio for mfg & trade:wholesaler;inventory/sales(87$)(s.a.)
ivsrrq	1959:01-1998:12 2	ratio for mfg & trade:retail trade;inventory/sales(87$)(s.a.)
pmnv	1959:01-1998:12 1	napm inventories index (percent)

Orders and unfilled orders

pmno	1959:01-1998:12 1	napm new orders index (percent)
pmdel	1959:01-1998:12 1	napm vendor deliveries index (percent)
mocmq	1959:01-1998:12 5	new orders (net)-consumer goods & materials, 1992 dollars (bci)
mdoq	1959:01-1998:12 5	new orders, durable goods industries, 1992 dollars (bci)
msondq	1959:01-1998:12 5	new orders, nondefense capital goods, in 1992 dollars (bci)
mo	1959:01-1998:12 5	mfg new orders: all manufacturing industries, total (mil$,sa)
mowu	1959:01-1998:12 5	mfg new orders: mfg industries with unfilled orders(mil$,sa)
mdo	1959:01-1998:12 5	mfg new orders: durable goods industries, total (mil$,sa)
mduwu	1959:01-1998:12 5	mfg new orders:durable goods indust with unfilled orders(mil$,sa)
mno	1959:01-1998:12 5	mfg new orders: nondurable goods industries, total (mil$,sa)
mnou	1959:01-1998:12 5	mfg new orders: nondurable gds ind.with unfilled orders(mil$,sa)
mu	1959:01-1998:12 5	mfg unfilled orders: all manufacturing industries, total (mil$,sa)
mdu	1959:01-1998:12 5	mfg unfilled orders: durable goods industries, total (mil$,sa)
mnu	1959:01-1998:12 5	mfg unfilled orders: nondurable goods industries, total (mil$,sa)
mpcon	1959:01-1998:12 5	contracts & orders for plant & equipment (bil$,sa)
mpconq	1959:01-1998:12 5	contracts & orders for plant & equipment in 1992 dollars (bci)

Stock prices

fsncom	1959:01-1998:12 5	NYSE common stock price index: composite (12/31/65=50)
fspcom	1959:01-1998:12 5	S&P's common stock price index: composite (1941-43=10)
fspin	1959:01-1998:12 5	S&P's common stock price index: industrials (1941-43=10)
fspcap	1959:01-1998:12 5	S&P's common stock price index: capital goods (1941-43=10)
fsput	1959:01-1998:12 5	S&P's common stock price index: utilities (1941-43=10)
fsdxp	1959:01-1998:12 1	S&P's composite common stock: dividend yield (% per annum)
fspxe	1959:01-1998:12 1	S&P's composite common stock: price-earnings ratio (%,nsa)

Exchange rates

exrus	1959:01-1998:12 5	United States effective exchange rate (merm)(index no.)
ex*rG*er	1959:01-1998:12 5	foreign exchange rate: Germany (deutsche mark per U.S.$)
exrsw	1959:01-1998:12 5	foreign exchange rate: Switzerland (swiss franc per U.S.$)
exrjan	1959:01-1998:12 5	foreign exchange rate: Japan (yen per U.S.$)
exrcan	1959:01-1998:12 5	foreign exchange rate: Canada (canadian $ per U.S.$)

Interest rates

fygt5	1959:01-1998:12 2	interest rate: U.S.treasury const maturities,5-yr.(% per ann,nsa)
fygt10	1959:01-1998:12 2	interest rate: U.S.treasury const maturities,10-yr.(% per ann,nsa)
fyaaac	1959:01-1998:12 2	bond yield: moody's aaa corporate (% per annum)
fybaac	1959:01-1998:12 2	bond yield: moody's baa corporate (% per annum)
fyfha	1959:01-1998:12 2	secondary market yields on fha mortgages (% per annum)
sfycp	1959:01-1998:12 1	spread fycp - fyff
sfygm3	1959:01-1998:12 1	spread fygm3 - fyff
sfygm6	1959:01-1998:12 1	spread fygm6 - fyff
sfygt1	1959:01-1998:12 1	spread fygt1 - fyff
sfygt5	1959:01-1998:12 1	spread fygt5 - fyff
sfygt10	1959:01-1998:12 1	spread fygt10 - fyff
sfyaaac	1959:01-1998:12 1	spread fyaaac - fyff
sfybaac	1959:01-1998:12 1	spread fybaac - fyff
sfyfha	1959:01-1998:12 1	spread fyfha - fyff

Money and credit quantity aggregates

fm1	1959:01-1998:12 6	money stock: m1(curr,trav.cks,dem dep,other ck'able dep)(bil$,sa)
fm2	1959:01-1998:12 6	money stock:m2(m1+o'nite rps,euro$,g/p&b/d mmmfs&sav&sm time dep(bil$,nsa)
fm3	1959:01-1998:12 6	money stock: m3(m2+lg time dep,term rp's&inst only mmmfs)(bil$,sa)

```
fm2dq    1959:01-1998:12 5    money supply-m2 in 1992 dollars (bci)
fmfba    1959:01-1998:12 6    monetary base, adj for reserve requirement changes(mil$,sa)
fmrra    1959:01-1998:12 6    depository inst reserves:total,adj for reserve req chgs(mil$,sa)
fmrnbc   1959:01-1998:12 6    depository inst reserves:nonborrow+ext cr,adj res req cgs(mil$,sa)
```

Price indexes

```
pmcp     1959:01-1998:12 1    napm commodity prices index (percent)
pwfsa    1959:01-1998:12 6    producer price index: finished goods (82=100,sa)
pwfcsa   1959:01-1998:12 6    producer price index:finished consumer goods (82=100,sa)
psm99q   1959:01-1998:12 6    index of sensitive materials prices (1990=100)(bci-99a)
punew    1959:01-1998:12 6    cpi-u: all items (82-84=100,sa)
pu83     1959:01-1998:12 6    cpi-u: apparel & upkeep (82-84=100,sa)
pu84     1959:01-1998:12 6    cpi-u: transportation (82-84=100,sa)
pu85     1959:01-1998:12 6    cpi-u: medical care (82-84=100,sa)
puc      1959:01-1998:12 6    cpi-u: commodities (82-84=100,sa)
pucd     1959:01-1998:12 6    cpi-u: durables (82-84=100,sa)
pus      1959:01-1998:12 6    cpi-u: services (82-84=100,sa)
puxf     1959:01-1998:12 6    cpi-u: all items less food (82-84=100,sa)
puxhs    1959:01-1998:12 6    cpi-u: all items less shelter (82-84=100,sa)
puxm     1959:01-1998:12 6    cpi-u: all items less midical care (82-84=100,sa)
gmdc     1959:01-1998:12 6    pce,impl pr defl:pce (1987=100)
gmdcd    1959:01-1998:12 6    pce,impl pr defl:pce; durables (1987=100)
gmdcn    1959:01-1998:12 6    pce,impl pr defl:pce; nondurables (1987=100)
gmdcs    1959:01-1998:12 6    pce,impl pr defl:pce; services (1987=100)
```

Average hourly earnings

```
lehcc    1959:01-1998:12 6    avg hr earnings of constr wkrs: construction ($,sa)
lehm     1959:01-1998:12 6    avg hr earnings of prod wkrs: manufacturing ($,sa)
```

Miscellaneous

```
hhsntn   1959:01-1998:12 1    u. of mich. index of consumer expectations(bcd-83)
```