

# Bayes Estimates in Multivariate Semiparametric Linear Models

Olaf Bunke

Humboldt University, Berlin

**Summary:** Bayes estimates are derived in multivariate linear models with unknown distribution. The prior distribution is defined using a Dirichlet prior for the unknown error distribution and a normal-Wishart distribution for the parameters. The posterior distribution for the parameters is determined and is a mixture of normal-Wishart distributions. The posterior mean of the observation distributions is a mixture of generalized Student distributions and of kernel estimates and empirical distributions based on "pseudoobservations". Explicit expressions are given in the special cases of location - scale and two-sample models. The calculation of "selfinformative" limits of Bayes estimates yields standard estimates.

**AMS 1980 subject classifications.** Primary 62F G05; Secondary 62A15.

**Key words and phrases.** Multivariate linear model, location-scale model, two-sample model, Dirichlet prior.

---

\* The research on this paper was carried out within the Sonderforschungsbereich 373 "Quantifikation und Simulation ökonomischer Prozesse" at Humboldt University Berlin and was printed using funds made available by the Deutsche Forschungsgemeinschaft.

# 1 Introduction

We consider  $n$  independent  $p$ -dimensional observations  $X_i$  with means and covariance matrices

$$(1.1) \quad E(X_i) = z_i B \quad , \quad D(X_i) = \Sigma \quad (i = 1, \dots, n),$$

where  $z_i$  are the rows of a known  $n \times k$  matrix  $Z$  of rank  $k$ . The  $k \times p$  matrix  $B$  and  $\Sigma$  are unknown parameters. The unknown distribution of  $X_i$  will be denoted by  $P_i$ . The matrix with rows  $X_i$  will be denoted by  $X$ . Bayes estimates are wellknown under normal distributions  $P_i$  and normal-Wishart prior distributions for the parameter

$$(1.2) \quad \xi = (B, \Omega) \quad , \quad \Omega = \Sigma^{-1}$$

(see De Groot (1970) or Zellner (1971)). Bayes estimates for  $B$  and  $\Sigma$  are not only useful, when some prior informations is incorporated into the statistical analysis in form of a prior distribution, but also as a basis for the construction of sensible estimates by adapting the parameters of the prior, e.g. estimating them (see Lindley (1972) and Efron and Morris (1973) or choosing their values by minimization of an estimate of the risk of the Bayes estimate (see Bunke (1986)). Moreover they are useful for producing by some limiting process the maximum likelihood estimates, namely by noninformative limits of priors and corresponding posteriors (see Box and Tiao (1973) and Hartigan (1983)). An alternative approach to the description of a noninformative situation could be to take the posterior as a new prior and calculate the corresponding posterior and repeate the procedure again and again hopefully converging to a value, which we will call “selfinformative limit“ (see section 3.). In parametric problems under some regularity conditions the posterior means calculated in this way converge to the maximum likelihood estimate (see Bunke, Hennig and Schmidt (1976)).

The model (1.1) is semiparametric, when the distributions  $P_i$  are unknown even under fixed parameters  $B, \Sigma$ . Our paper is concerned with the characterization of posterior distributions and corresponding Bayes estimates of  $B, \Sigma$  and of  $P_i$  and with the structure of their selfinformative limits in some special cases. This may give some insight into the problem of maximum likelihood estimation in semiparametric models, its treatment being still not complete (see Gill (1989)). The special case of Bayes estimation in a univariate semiparametric location model has already been treated under different priors by several authors (see Diaconis and Freedman (1986) and its discussion).

# 2 Prior and Posterior Distributions

The Dirichlet distribution  $D_\alpha$  of a random (probability) distribution  $G$  will be used to define our priors. It is determined by a finite measure  $\alpha$  on  $(R^p, \mathcal{B}^p)$ . The realizations of  $G$  are distributions on  $(R^p, \mathcal{B}^p)$ .  $D_\alpha$  is the distribution under which for

each finite partition of  $R^p$  into disjoint Borel sets  $B_1, \dots, B_m$  the random vector  $q := (G(B_1), \dots, G(B_m))$  with values in  $Q = \{q \in R^m \mid q_j \geq 0, \sum_j q_j = 1\}$  follows a Dirichlet distribution with density

$$(2.1) \quad f(q) \propto \prod_j \alpha(B_j)^{q_j-1}.$$

The mean of  $G$  under  $D_\alpha$  is

$$(2.2) \quad \int G D_\alpha(dG) = \beta = a^{-1}\alpha,$$

where  $a = \alpha(R^p)$ . Under  $D_\alpha$  almost all realizations  $G$  are distributions with means  $m_G$  and covariance matrices  $C_G$  (see Hartigan (1983), Ferguson (1973) and Simar (1984) for details).

We will introduce a prior for the unknown parameters and distributions which allows a calculation of the posterior distribution and generalizes the prior used in Diaconis and Freedman (1986) (first construction there). To define the prior we write

$$(2.3) \quad X_i = z_i A + U_i \Lambda^{-1/2},$$

and assume that

- (i) the vector  $\mathbb{1} = (1, \dots, 1)^t$  is contained in the linear space  $R(Z)$  generated by the columns of  $Z$ ,
- (ii) the  $p$ -dimensional random vectors  $U_1, \dots, U_n$  are i.i.d. with distribution  $G$  if  $A, \Lambda, G$  are fixed and
- (iii) under the prior the parameter  $\vartheta = (A, \Lambda)$  and  $G$  are independent,  $G$  having a Dirichlet distribution  $D_\alpha$  with  $\beta = a^{-1}\alpha = N(0, I)$  and  $\vartheta$  having a normal-Wishart density with  $f = r + p - k + 1$  degrees of freedom:

$$(2.4) \quad f(\vartheta) \propto |\Lambda|^{r/2} e[\Lambda(S_0 + (A - A_0)^t \Gamma_0 (A - A_0))],$$

where  $r > k$ ,  $e[M] = \exp[-trM/2]$ ,  $A_0$  is a  $k \times p$  matrix and  $\Gamma_0, S_0$  are positive definite symmetric matrices.

The equation (2.3) generating the observations  $X_i$  determines the relationship between (i) the actually interesting parameters  $B, \Sigma$  in (1.1) and distributions  $P_i = P_i(A, \Lambda, G)$  of  $X_i$  and (ii) the parameters  $A, \Lambda$  and distribution  $G$ . It holds

$$(2.5) \quad B = A + Z^+ \mathbf{1}_{m_G} \Lambda^{-1/2}, \quad \Sigma = \Lambda^{-1/2} C_G \Lambda^{-1/2}$$

where  $Z^+ = (Z'Z)^{-1}Z'$ .

It is important to know prior means of the parameters of interest to allow a sensible choice of the constants in (2.4).

**Theorem 2.1** *Under the prior introduced for  $\vartheta, G$  and (1.1), (2.3) it holds, that  $E(G) = N(0, I)$  and*

$$(2.6) \quad E(B) = A_0 \quad , \quad E(\Sigma) = a(r - k)^{-1}(a + 1)^{-1}S_0.$$

We now discuss the posterior distribution of  $\vartheta, G$ , which determines the posterior means of  $B, \Sigma$  by (2.5) and of the unknown distributions  $P_i$  by (2.3). Because under the condition of fixed  $A, \Lambda, G$  the random vectors  $Y_i = (X_i - z_i A)\Lambda^{1/2}$  are i.i.d. with distribution  $G$ , we obtain the conditional posterior of  $G$  under the condition of fixed  $A, \Lambda$  as the Dirichlet  $D_{\alpha_Y}$  corresponding to

$$(2.7) \quad \alpha_Y = \alpha + \sum_i \delta_{Y_i}$$

where  $\delta_y$  (see Ferguson (1973)) denotes the Dirac measure degenerated at  $y$ . Therefore (see (2.3)) the posterior mean of  $P_i = P_i(A, \Lambda, G)$  (the distribution of  $X_i$ ) will be (for  $P^X$ -almost all  $X$  w.r.t. the marginal distribution  $P^X$  of  $X$ )

$$(2.8) \quad \tilde{P}_i = E(P_i|X) = \kappa E_X N(z_i A, \Lambda^{-1}) + \bar{\kappa} E_X Q_{\tilde{Y}},$$

where

$$(2.9) \quad \tilde{Y}_j = Y_j \Lambda^{-1/2} + z_i A = X_j + (z_i - z_j)A,$$

$$(2.10) \quad \kappa = a/(a + n) \quad , \quad \bar{\kappa} = 1 - \kappa$$

and  $Q_{\tilde{Y}}$  denotes the empirical distribution of  $\tilde{Y}_1, \dots, \tilde{Y}_n$ ,  $E_X$  denoting the mean w.r.t. the posterior distribution of  $\vartheta$ .

The posterior distribution for  $\vartheta$  is determined by the prior of  $\vartheta$  and the marginal distribution  $W$  of the matrix  $U$  with rows  $U_i$  (see (2.3)). Under  $W$  the conditional distribution of  $U_m$  under the condition of fixed  $U_1, \dots, U_{m-1}$  is

$$(2.11) \quad (a + m - 1)^{-1} \left[ aN(0, I) + \sum_{i=1}^{m-1} \delta_{U_i} \right]$$

(see Hartigan (1983)) and therefore the probability is positive, that some of the rows  $U_i$  are identical. More specifically it holds (see Ghorai and Rubin (1982) and Lo (1984))

$$(2.12) \quad W = \sum_{c \in C} W(M_c) W_c,$$

where  $C$  is the set of all partitions

$$(2.13) \quad c : \quad N = \{1, \dots, n\} = \sum_{t=1}^{|c|} N_{ct}$$

of the set  $N$  into disjoint subsets  $N_{ct}$ ,  $L_c$  is the linear subspace of all  $n \times p$  matrices  $U$  with identical rows  $U_i = U_j$  if  $i, j \in N_{ct}$ ,

$$(2.14) \quad M_c = \{U \in L_c \mid U_i \neq U_j \text{ if } i \in N_{ct}, j \in N_{ch}, t \neq h\}$$

and  $W_c$  is the conditional distribution of  $U$  under the condition  $U \in M_c$ . From (2.11) it easily proven by induction, that

$$(2.15) \quad \pi_c = W(M_c) = n_c a^{|c|-1} [(a+1) \cdot (a+n-1)]^{-1},$$

$$(2.16) \quad n_c = \prod_{t=1}^{|c|} |N_{ct}|! \quad , \quad (|N_{ct}| : \text{ number of elements in } N_{ct}).$$

$\tilde{W}_c$  is the distribution of a random matrix  $U$  with values in  $L_c$ , the different rows being i.i.d. with distribution  $N(0, I)$ , and  $W_c$  is the restriction of  $\tilde{W}_c$  on  $M_c$ . A consequence of the structure (2.12) is the mixture structure of the posterior distribution  $P_X$  of  $\vartheta$

$$(2.17) \quad P_X = \sum_{c \in C} P(U \in M_c \mid X) P_{c,X} \quad ,$$

where  $P_{c,X}$  is the posterior of  $\vartheta$  under a model (2.3) with a fixed distribution  $W_c$  of  $U$ . Fortunately, some or even many of the terms in (2.17) may vanish together with the posterior probabilities for  $U \in M_c$ , as shown in the next theorem 2.2. For this we introduce some notation:

The image of the mapping (2.3) under the restriction  $U \in M$  will be

$$(2.18) \quad T(M) := \{X = ZB + U \mid B \in R^{k \times p}, U \in M\}.$$

We introduce the following subsets of partitions:

$$(2.19) \quad D(c) = \{\underline{c} \in C \mid T(L_{\underline{c}}) = T(L_c)\}$$

$$(2.20) \quad D = \{c \in C \mid T(L_c) = R^{n \times p}\}$$

and (denoting the dimension of  $T(L_c)$  by  $d_c$ ) the subsets of observations

$$(2.21) \quad T_c = \{X \in T(M_c) \mid X \notin T(L_{\underline{c}}) \quad \forall \underline{c} : d_{\underline{c}} \leq d_c, T(L_{\underline{c}}) \neq T(L_c)\},$$

$$(2.22) \quad T(c) = \bigcup_{\underline{c} \in D(c)} T_{\underline{c}} \quad , \quad T = \bigcup_{c \in D} T_c.$$

The set  $C$  of partitions is obviously the union of equivalence classes  $D(c)$  ( $= D(\underline{c})$  for  $\underline{c} \in D(c)$ ) of partitions leading to same images  $T(L_c)$ . The set  $T(c)$  of observations corresponds just to the class  $D(c)$ , including only observations generated by “errors“  $U$  from  $M_c$  ( $R^{n \times p}$  has been already partitioned in the disjoint subsets  $M_c$ ) and leaving out the observations belonging to images  $T(L_{\underline{c}})$  of smaller dimension (or different from  $T(L_c)$ , if the dimension is the same).  $D$  and  $T$  correspond to the full dimension  $d_c = n$ . We denote by  $\lambda_L$  the Lebesgue measure on the  $\sigma$ -algebra of Borel sets of the linear subspace  $L \subset R^{n \times p}$  and by  $\lambda_L$  its extension to  $\mathcal{B}^{n \times p}$ . Then it is easy to see, that  $T(L_c)$ ,  $T(M_c)$  and  $T(c)$  differ only by a nullset w.r.t. the Lebesgue measure  $\lambda_c := \lambda_{T(L_c)}$ .

In the following we will use the positive semidefinite (possibly singular) matrix

$$(2.23) \quad \Omega_c = Z\Gamma_0^{-1}Z^t + K_c,$$

the elements  $k_{cij}$  of  $K_c$  being equal to one, if  $i$  and  $j$  belong to the same subset  $N_{ct}$  and zero otherwise, and the constants

$$(2.24) \quad d_c = \text{rank } \Omega_c = \dim T(L_c), \quad s_c = r + d_c + p + 1 - k.$$

Finally we denote the determinant of a matrix  $G$  by  $|G|$ , the Moore-Penrose inverse of  $G$  by  $G^+$  and we introduce the functions

$$(2.25) \quad \tilde{w}_c(X) = |\Omega_c^+|^{p/2} |S_0 + (X - ZA_0)^t \Omega_c^+ (X - ZA_0)|^{-s_c/2}$$

$$(2.26) \quad w_c(X) = g_c^{-1} \tilde{w}_c(X).$$

The constants

$$(2.27) \quad g_c = \int \tilde{w}_c(X) d\lambda_c(X) = g(s_c)$$

may be shown to depend on  $c$  through  $s_c$  writing  $w_c(X)$  as the density of a generalized Student distribution (see De Groot (1979)) in the matrix variable  $Y = U_c X$  defined by the diagonalization or spectral decomposition  $\Omega_c = U_c^t \text{Diag}[w_{c1}, \dots] U_c$ .

**Theorem 2.2** *Under the above assumptions, for all  $\bar{c} \in C$  and for  $P^X$  almost all  $X$  in  $T(\bar{c})$  it holds, that*

$$(2.28) \quad P(U \in M_c | X) = \pi_c w_c(X) / \sum_{\underline{c} \in D(\bar{c})} \pi_{\underline{c}} w_{\underline{c}}(x),$$

if  $c \in D(\bar{c})$  and

$$(2.29) \quad P(U \in M_c | X) = 0 \quad \text{if} \quad c \in C - D(\bar{c}).$$

Let  $\nu_p$  the Lebesgue measure on the  $\sigma$ -algebra of Borel sets of the set of positive definite  $p \times p$  matrices. An explicit expression for the density of  $P_{c,X}$  w.r.t. a product measure of the form  $\mu_c = \lambda_{R_c} \times \nu_p$  can be derived for  $P^X$  - almost all  $X$  in  $T(c)$ . The linear space  $R_c$  will be generated by the matrix  $\Delta_c$  defined in (2.34):

$$(2.30) \quad R_c = \{\Delta_c H \mid H \in R^{k \times p}\}.$$

**Theorem 2.3** *Under the above assumptions and for  $P^X$ -almost all  $X \in T(L_c)$  the distributions  $P_{c,X}$  have the densities w.r.t.  $\mu_c$*

$$(2.31) \quad p_c(\vartheta | X) \propto |\Lambda|^{(r+d_c)/2} e[\Lambda\{S_c + (A - \tilde{A}_c)^t \Delta_c^+ (A - \tilde{A}_c)\}],$$

where

$$(2.32) \quad S_c = S_o + (X - Z A_0)^t \Omega_c^+ (X - Z A_0),$$

$$(2.33) \quad \tilde{A}_c = A_0 + \Gamma_0^{-1} Z^t \Omega_c^+ (X - Z A_0),$$

$$(2.34) \quad \Delta_c = \Gamma_0^{-1} - \Gamma_0^{-1} Z^t \Omega_c^+ Z \Gamma_0^{-1}.$$

The means of  $A$  and  $\Lambda^{-1}$  under the “generalized normal-Wishart” density (2.31) are  $\tilde{A}_c$  and (see De Groot (1970))

$$(2.35) \quad \widetilde{\Lambda_c^{-1}} = (d_c + r - k)^{-1} S_c.$$

These means yield together with (2.5), (2.17) and theorem 2.2 the posterior means of  $B$  and  $\Sigma$ , which are Bayes estimates w.r.t. quadratic risks. For each observation  $X \in T_{\bar{c}}$  the terms appearing in the expressions correspond to partitiones  $c$  leading to the same linear space  $T(L_c) = T(L_{\bar{c}})$ :

**Theorem 2.4** *Under the above assumptions and for  $P^X$ - almost all  $X \in T(L_{\bar{c}})$  ( $\bar{c} \in C$ ) leading to densities (2.31) with some matrices  $\tilde{A}_c, S_c$  for all  $c \in D(\bar{c})$  the posterior means  $\tilde{B}$  and  $\tilde{\Sigma}$  of  $B$  and  $\Sigma$  are*

$$(2.36) \quad \tilde{B} = \sum_{c \in D(\bar{c})} r_c \tilde{B}_c, \quad \tilde{\Sigma} = \sum_{c \in D(\bar{c})} r_c \tilde{\Sigma}_c.$$

where  $r_c = P\{U \in M_c \mid X\}$  and

$$(2.37) \quad \tilde{B}_c = \tilde{A}_c + \bar{\kappa} Z^+ \mathbb{1}(\bar{x} - \bar{z} \tilde{A}_c),$$

$$(2.38) \quad \bar{x} = n^{-1} \sum_i X_i, \quad \bar{z} = n^{-1} \sum_i z_i,$$

$$(2.39) \quad \tilde{\Sigma}_c = \omega[\kappa \widetilde{\Lambda_c^{-1}} + \bar{\kappa}(q_c \widetilde{\Lambda_c^{-1}} + R_c - \bar{\kappa} \bar{r}_c^t \bar{r}_c)],$$

where

$$(2.40) \quad \omega = (a + n)/(a + n + 1),$$



$$(2.41) \quad q_c = \text{tr}[\Delta_c(n^{-1}Z^tZ - \bar{\kappa}\bar{z}^t\bar{z})],$$

$$(2.42) \quad r_{ci} = X_i - z_i\tilde{A}_c \quad , \quad \bar{r}_c = n^{-1} \sum_i r_{ci},$$

$$(2.43) \quad R_c = n^{-1} \sum_i r_{ci}^t r_{ci}.$$

Moreover the posterior mean of  $P_i$  is a mixture

$$(2.44) \quad \tilde{P}_i = \sum_{c \in D(\bar{c})} r_c [\kappa F_{i,c} + \bar{\kappa}(\nu_{ic}P_{ic} + \bar{\nu}_{ic}\hat{F}_{ic})]$$

of distributions, where  $F_{i,c}$  has the density

$$(2.45) \quad f_{i,c}(y) = f(y \mid s_c + 1, h_{ic} + 1, S_c, z_i\tilde{A}_c),$$

$$(2.46) \quad h_{ic} = z_i\Delta_c z_i^t \quad ,$$

$$(2.47) \quad f(y \mid a, b, S, d) \propto |S + b^{-1}(y - d)^t(y - d)|^{-a/2}$$

being the density of a generalized Student distribution,  $P_{ic}$  has the density

$$(2.48) \quad p_{ic}(y) = \sum_{j \in J_{ic}} m_{ic}^{-1} h_{ijc}^{-p} q_{ic}([y - X_i(j, c)]/h_{ijc})$$

being a “kernel desity estimate“ depending on “pseudoobservations“

$$(2.49) \quad X_i(j, c) = z_i\tilde{A}_c + (X_j - z_j\tilde{A}_c),$$

and  $\hat{F}_{ic}$  is the empirical distribution of the pseudo observations  $X_i(j, c), j \in V_{ic}$ , using the notation

$$(2.50) \quad h_{ijc}^2 = (z_i - z_j) \Delta_c (z_i - z_j)^t,$$

$$(2.51) \quad J_{ic} = \{j \in N \mid h_{ijc} > 0\} \quad , \quad V_{ic} = N - J_{ic},$$

$$(2.52) \quad q_{ic}(y) = f(y \mid s_c, 1, S_c, 0),$$

$$(2.53) \quad \nu_{ic} = m_{ic}/n \quad , \quad \bar{\nu}_{ic} = 1 - \nu_{ic} \quad , \quad m_{ic} = |J_{ic}|.$$

### 3 The Location-Scale Model

The location scale model is the special case  $k = 1$ ,  $Z = \mathbb{1}$  of (1.1), where we write  $\gamma = \Gamma_0$ . Here we have  $T(M_c) = M_c$ , these sets being different for different  $c$ . Therefore  $D(c) = \{c\}$  for all  $c$  and

$$(3.1) \quad P(U \in M_{\underline{c}}, \mid X) = \delta_{c \underline{c}}$$

for  $P^X$ -almost all  $X \in M_c$ , so that for these  $X$  the Bayes estimates of  $B$  and  $\Sigma$  are of the form (2.37) and (2.39), provided that the distribution  $P_{c,X}$  has a density of the form (2.31), see theorem 2.4. This is the case, because following from the definition of  $W_c$  the distribution  $P_{c,X}$  is the posterior under the parametric model (2.3), where  $i = i_1, \dots, i_{|c|}$  and where  $i_t$  is a fixed index from  $N_{ct}$  for each  $t$ , the random vectors  $U_{i_1}, \dots, U_{i_c}$  being i.i.d. with distribution  $N(0, I)$ . In this normal location-scale model the posterior density is wellknown (see Zellner (1971)) and of the form (2.31) with

$$(3.2) \quad S_c = S_0 + |c| H_c + |c| \lambda_c (\bar{x}_c - \tilde{A}_c),$$

$$(3.3) \quad \bar{x}_c = \sum_t X_{i_t} / |c| ,$$

$$(3.4) \quad H_c = \sum_t (X_{i_t} - \bar{x}_c)^t (X_{i_t} - \bar{x}_c) / |c| ,$$

$$(3.5) \quad \lambda_c = \gamma(\gamma + |c|)^{-1},$$

$$(3.6) \quad \tilde{A}_c = \lambda_c A_0 + (1 - \lambda_c) \bar{x}_c.$$

For  $P^X$ -almost all  $X \in M_c$  we obtain the Bayes estimates

$$(3.7) \quad \tilde{B}_c = \kappa[\lambda_c A_0 + (1 - \lambda_c) \bar{x}_c] + \bar{\kappa} \bar{x},$$

$$(3.8) \quad \tilde{\Sigma}_c = \omega[\kappa(1 + \bar{\kappa}[\gamma + |c|]^{-1})\widetilde{\Lambda_c^{-1}} + \bar{\kappa} H_c + \kappa \bar{\kappa}(\bar{x} - \tilde{A}_c)^t(\bar{x} - \tilde{A}_c)]$$

where

$$(3.9) \quad \bar{x} = n^{-1} \sum_{i=1}^n x_i, \quad \widetilde{\Lambda_c^{-1}} = (|c| + r - 1)^{-1} S_c.$$

We remark, that  $\bar{x}_c$  and  $H_c$  are the sample mean and covariance matrix of the subsample consisting of the different observations  $X_i$ .

The Bayes estimate  $\tilde{P}_i$  of  $P_i$  will be (see (2.44))

$$(3.10) \quad \tilde{P}_c = \kappa F_c + \bar{\kappa} \hat{F},$$

where  $F_c$  has the density  $f(y | s_c + 1, (\gamma + |c|)^{-1} + 1, S_c, B_c)$  and  $\hat{F}$  is the empirical distribution of the observations  $x_1, \dots, x_n$ , all constants (2.46) vanishing.

Remark 1: Because of the law of large numbers, for a fixed distribution  $G$  and fixed parameters  $B$  and  $\Sigma$ , the mean  $\bar{x}_c$  and  $H_c$  will be consistent estimates of  $B$  and  $\Sigma$  resp. and the empirical distribution  $\hat{F}$  will be consistent for the distribution  $P$  of  $X_i$ . Therefore all our Bayes estimates  $\tilde{B}, \tilde{\Sigma}, \tilde{P}$  will be consistent in the frequentist sense.

Remark 2: We will investigate “selfinformative“ limits of Bayes estimates, which are obtained by taking the posterior  $P_X^\vartheta$  in place of the original prior for  $\vartheta$  in a parametric problem, calculating the corresponding posterior and repeating this procedure again and again. . We will define a selfinformative limit in a precise manner, which is equivalent to the above mentioned iterative procedure if the determination of the conditional expectation given by (2.36), (2.39), (2.44) is used.

We consider  $m$  replications of the model (1.1), that is precisely, we take  $\tilde{n} = mn$  in place of  $n$ ,  $\tilde{Z} = \mathbb{1}_m \otimes Z$  in place of  $Z$  and take independent observations  $\tilde{X}_i (i = 1, \dots, \tilde{n})$  in place of  $X_i$ ,  $\otimes$  denoting the Kronecker product. Now we have the model

$$(3.11) \quad E\tilde{X} = \tilde{Z}B, \quad D\tilde{X} = \Sigma \otimes I_{\tilde{n}}$$

replacing (1.1).

**Definition 3.1** *We assume  ${}_mX = \mathbb{1}_m \otimes X$  and denote by  $\tilde{B}, \tilde{\Sigma}, \tilde{P}_i$  the Bayes estimates corresponding to observations  $\tilde{X} \in R^{\tilde{n} \times p}$  defined by (2.36), (2.37), (2.39) and (2.44). The limit of  $\tilde{B}, \tilde{\Sigma}$  or of  $\tilde{P}_i$  calculated for  $\tilde{X} = {}_mX$  and for  $m \rightarrow \infty$ , if it exists, is called a selfinformative limit.*

We calculate now selfinformative limits in the location-scale model and assume, that all observations  $X_i$  are distinct. Then the Bayes estimates are (3.7), (3.8) and (3.10), where  $\bar{x}_c$  and  $H_c$  are the mean  $\bar{x}$  and the covariance matrix

$$(3.12) \quad S = n^{-1} \sum_{i=1}^n (X_i - \bar{x})^t (X_i - \bar{x})$$

of the whole sample  $X_1, \dots, X_n$  resp. The “sample“  ${}_mX$  of size  $\tilde{n}$  has the same mean and covariance matrix and also the same empirical distribution (each value  $X_i$  appears  $m$  times and its probability under the empirical distribution is  $m/\tilde{n} = 1/n$ ). The matrices  $\tilde{A}_c$  and  $\tilde{S}_c$  for this sample have the limits (using  $|c| \tilde{n}$ )

$$(3.13) \quad \lim_{m \rightarrow \infty} \tilde{a}_c = \bar{x}, \quad \lim_{m \rightarrow \infty} m^{-1} \tilde{S}_c = nS,$$

so that

$$(3.14) \quad \widetilde{\Lambda_c^{-1}} = (r + \tilde{n} - k)^{-1} \tilde{S}_c \rightarrow S.$$

We obtain the selfinformative limits

$$(3.15) \quad \tilde{B} \rightarrow \bar{x}, \quad \tilde{\Sigma} \rightarrow S, \quad \tilde{P} \rightarrow \hat{P}$$

because of (3.7), (3.8). (3.10) and

$$(3.16) \quad \lambda_c = \gamma(\gamma + mn)^{-1} \rightarrow 0, \quad \kappa = a(a + mn)^{-1} \rightarrow 0.$$

## 4 The Two-Sample Model

The two-sample model is the special case  $k = 2$ ,

$$(4.1) \quad Z = \begin{pmatrix} \mathbb{1}_{n_1} & 0 \\ 0 & \mathbb{1}_{n_2} \end{pmatrix}, \quad B = \begin{pmatrix} b_1^t \\ b_2^t \end{pmatrix}$$

of (1.1), the size of the  $i$ -th sample being  $n_i$ . We will restrict our attention to the estimation of the means  $b_1$  and  $b_2$  of both samples and assume

$$(4.2) \quad \Gamma_0 = \begin{pmatrix} w_1 & 0 \\ 0 & w_2 \end{pmatrix}, \quad A_0 = \begin{pmatrix} a_{01}^t \\ a_{02}^t \end{pmatrix}.$$

It is easy to see, that  $T(L_c) = R^{n \times p}$  for the partitions

$$(4.3) \quad c_* : N = \sum_{i=1}^n \{i\}$$

and, (for all  $i, j$  with  $1 \leq i \leq n_1 < j \leq n = n_1 + n_2$ )

$$(4.4) \quad c(i, j) : N = \sum_{m \neq i, j} \{m\} + \{i, j\},$$

while  $T(L_c)$  is a subspace and Lebesgue null set of  $R^{n \times p}$  for all other partitions. Therefore we have

$$(4.5) \quad D = \{c_*\} + V, \quad V = \{c(i, j) \mid i \leq n_1 < j \leq n\}$$

(see (2.19)). Then the set  $T$  (see (2.22)) is the complement of the Lebesgue null set

$$(4.6) \quad \overline{T} = \bigcup_{1 \leq i < k \leq n_1} L_{ik} \cup \bigcup_{n_1 < j < m \leq n} L_{jm} \cup \bigcup_{\substack{1 \leq i < k \leq n_1 : i \neq k \\ n_1 < j < m \leq n : j \neq m}} L_{ikjm},$$

being the union of linear subspaces

$$(4.7) \quad L_{ik} = \{X \in R^{n \times p} \mid X_i = X_k\},$$

$$(4.8) \quad L_{ikjm} = \{X = ZA + U \mid A \in R^{2 \times p}, U \in R^{n \times p} : U_i = U_j, U_k = U_m\}.$$

Theorem 2.4 gives after a cumbersome algebraic calculation (see appendix):

**Theorem 4.1** *The posterior means of  $b_1, b_2$  in the semiparametric two-sample model (1.1), (4.1) under the above assumptions on the prior and (4.2) are given for  $P^X$ -almost all  $X \in T$  by*

$$(4.9) \quad \tilde{B} = r_{c_*} \tilde{B}_{c_*} + \sum_{1 \leq i \leq n_1 < j \leq n} r_{c(i,j)} \tilde{B}_{c(i,j)},$$

$$(4.10) \quad \tilde{B}_c = \tilde{A}_c + \bar{\kappa} \mathbb{1} \left( \bar{x} - \frac{n_1}{n} \tilde{a}_{c1}^t - \frac{n_2}{n} \tilde{a}_{c2}^t \right),$$

$$(4.11) \quad \tilde{a}_{c_*m}^t = (n_m + w_m)^{-1} (n_m \tilde{x}_m + w_m a_{0m}^t), \quad (m = 1, 2),$$

$$(4.12) \quad \tilde{a}_{c(i,j)1}^t = d[\mathbb{1}'X + \lambda_2(X_i - X_j) + X_i + (1 - n_1 - \lambda_2)a_{01}^t + (\lambda_2 - n_2)a_{02}^t],$$

$$(4.13) \quad \tilde{a}_{c(i,j)2}^t = d[\mathbb{1}'X + \lambda_1(X_j - X_i) + X_j + (\lambda_1 - n_1)a_{01}^t + (1 - n_2 - \lambda_1)a_{02}^t],$$

where

$$(4.14) \quad \bar{x}_1 = n_1^{-1} \sum_{n=1}^{n_1} X_i, \quad \bar{x}_2 = n_2^{-1} \sum_{i=n_1+1}^n X_i,$$

$$(4.15) \quad \lambda_m = n_m - 1 + w_m \quad (m = 1, 2),$$

$$(4.16) \quad d = (n - 1 + w_1 + w_2)^{-1}.$$

The weights  $r_c$  in (4.9) are determined by:

$$(4.17) \quad r_c = s_c / (s_{c_*} + \sum_{1 \leq i \leq n_1 < j \leq n} s_{c(i,j)}),$$

where

$$(4.18) \quad s_{c_*} = [(n_1 w_1^{-1} + 1)(n_2 w_2^{-1} + 1)]^{-p/2} |S_0 + Q_*|^{-(g+n)/2},$$

$$(4.19) \quad Q_* = \tilde{X}^t \tilde{X} - \sum_{m=1}^2 (n_m + w_m)^{-1} n_m^2 \tilde{x}_m^t \tilde{x}_m,$$

where

$$(4.20) \quad \tilde{X} = X - Z A_0,$$

$$(4.21) \quad \tilde{x}_1 = n_1^{-1} \sum_{i=1}^{n_1} \tilde{X}_i, \quad \tilde{x}_2 = n_2^{-1} \sum_{i=n_1+1}^{n_2} \tilde{X}_i,$$

$$(4.22) \quad g = r + p + 1 - k,$$

and

$$(4.23) \quad s_{c(i,j)} = (d^{-1} w_1 w_2)^{p/2} |S_0 + Q_{i,j}|^{-(g+n)/2},$$

where

$$(4.24) \quad Q_{i,j} = \tilde{X}^t \tilde{X} - d^{-1} [T_{ij}^t T_{ij} + W_{ij} - H_{ij}],$$

$$(4.25) \quad T_{ij} = \sum_{m=1}^n \tilde{X}_m - \tilde{X}_i - \tilde{X}_j,$$

$$(4.26) \quad W_{ij} = \tilde{X}_i^t T_i + T_i^t \tilde{X}_i + \tilde{X}_j^t T_j + T_j^t \tilde{X}_j + (d - \lambda_1) \tilde{X}_i^t \tilde{X}_i + (d - \lambda_2) \tilde{X}_j^t \tilde{X}_j$$

$$(4.27) \quad H_{ij} = \lambda_1 \lambda_2 \tilde{X}_{i,j}^t \tilde{X}_{i,j} + \tilde{X}_{i,j}^t (\lambda_1 T_j - \lambda_2 T_i) + (\lambda_1 T_j - \lambda_2 T_i)^t \tilde{X}_{i,j},$$

$$(4.28) \quad T_i = \sum_{m=1}^{n_1} \tilde{X}_m - \tilde{X}_i, \quad T_j = \sum_{m=n_1+1}^n \tilde{X}_m - \tilde{X}_j, \quad X_{i,j} = \tilde{X}_i - \tilde{X}_j.$$

The Bayes estimates  $\tilde{\Sigma}$  and  $\tilde{P}_i$  may be calculated in an analogous manner but yield even more complicated expressions and therefore are omitted here.

The calculation of the selfinformative limit of the Bayes estimate (4.9) of  $B$  yields the maximum likelihood estimate under normal distribution  $G = N(0, 1)$ :

**Theorem 4.2** *Under the assumptions of theorem 4.1 and for  $P^X$ -almost all  $X \in T$  the selfinformative limits of the Bayes estimates of the means  $b_1, b_2$  are*

$$(4.29) \quad \tilde{b}_m^t \rightarrow \bar{x}_m.$$

## References

- Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading.
- Bunke, H., Hennig, C. and Schmidt, W.H. (1973). Weighted combination of prior and sample information and parameter estimation. *Math. Operationsforsch. u. Statist.* **7**, 665-675.
- Bunke, O. (1986). Assessing the performance of regression estimators and models under nonstandard conditions. Seminarbericht Nr. 89, Humboldt-Universität zu Berlin, Sektion Mathematik.
- De Groot, M.H. (1970). *Optimal Statistical Decisions*. McGraw Hill, New York.
- Diaconis, P. and Friedman, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14**, 1-26.
- Efron, B. and Morris, C. (1973). Empirical Bayes on vector observations: An extension of Stein's method. *Biometrika* **59**, 335-347.
- Ferguson, T. (1973). Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209-230.
- Gill, R.D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part. 1). *Scand. J. Stat.* **16**, No.2, 97-128.
- Ghorai, J.K. and Rubin, H. (1982). Bayes risk consistency of nonparametric density estimates. *Australian J. Statist.*, 51 - 66.
- Hartigan, J.A. (1983). *Bayes Theory*. Springer, New York.
- Lindley, D.V. and Smith, A.F.M. (1972). Bayes estimates for the linear model (with discussion). *J. Roy. Statist. Soc., Ser. B*, **34**, 1-41.



Simar, L. (1984). A survey of Bayesian approaches to nonparametric statistics. *Statistics* **15**, 121 - 142.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. Wiley, New York.

## 5 Appendix

### 1. Proof of theorem 2.1:

(2.6) follows from (2.5), the prior independence of  $\vartheta$  and  $G$ , the means

$$(A.1) \quad E m_G = m_\beta = 0 \quad , \quad E C_G = a(a+1)^{-1} C_\beta = a(a+1)I$$

under a Dirichlet  $D_\alpha$  (see Ferguson (1973)) and the means

$$(A.2) \quad EA = A_0 \quad , \quad E\Lambda^{-1} = (r-k)^{-1}S$$

under (2.4) (see De Groot (1970)).

### 2. Proof of theorem 2.2:

Because of the definition of  $W_c$  and  $\Omega_c$  the conditional distribution of  $X$  under the conditions  $U \in M_c$  and of a fixed  $\Lambda$  is the normal  $N(ZA_0, \Lambda^{-1} \otimes \Omega_c)$ . It has the density

$$(A.3) \quad f(X | c, \Lambda) \propto |\Lambda|^{d_c/2} e[\Lambda(X - ZA_0)^t \Omega_c^+ (X - ZA_0)]$$

w.r.t.  $\lambda_c$ , as stated in the following lemma, which may easily be proven using principal components of  $X$  and of its subvector  $X_2$  and the known densities of normal distributions with nonsingular covariance matrices.

**Lemma 1:** *Assume*

$$X \sim N_n(0, \Sigma) \quad , \quad X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad , \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad ,$$

where  $X_2 \in R^k$  with  $k < n$  and  $\Sigma_{22} \in R^{k \times k}$ . We use the notation

$$(A.4) \quad L_1 = \{\Sigma_{11}b | b \in R^{n-k}\}, \quad L_2 = \{\Sigma_{22}b | b \in R^k\}$$

The distribution of  $X$  has w.r.t.  $\lambda_{L_1 \times L_2}$  the density  $p(x) \propto e[x^t \Sigma^+ x]$ . For almost all  $X_2 \in L_2$  the conditional distribution of  $X_1$  under the condition  $X_2$  has w.r.t.  $\lambda_{L_1}$  the density

$$(A.5) \quad p(X_1|X_2) \propto e[(X_1 + \Sigma_{12}\Sigma_{22}^+X_2)^t(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^+\Sigma_{21})^+(X_1 + \Sigma_{12}\Sigma_{22}^+X_2)].$$

Using the integral

$$(A.6) \quad \int |\Lambda|^b e[\Lambda S] d\Lambda = k_b |S|^{-(b+p+1)/2}$$

(see Zellner (1971)) we may integrate (A.3) w.r.t. the prior distribution of  $\Lambda$  (see (2.4)) and see, that the conditional distribution  $P_c$  of  $X$  under the condition  $U \in M_c$  has a density w.r.t.  $\lambda_{T(L_c)} =: \lambda_c$ , which is given by  $w_c(X)$  (see (2.26)). Now it is easy to see, that (see (2.21))

$$(A.7) \quad T(L_c) = T_c \cup \tilde{T}_c \quad \text{with some set } \tilde{T}_c \text{ obeying } \lambda_c(\tilde{T}_c) = 0,$$

$$(A.8) \quad \lambda_{\underline{c}}(T_c) \leq \lambda_{\underline{c}}(T(L_c)) = 0 \quad \text{if } d_{\underline{c}} > d_c$$

$$(A.9) \quad T_c \cap T(L_{\underline{c}}) = \emptyset \quad \text{if } d_{\underline{c}} \leq d_c, \quad T(L_{\underline{c}}) \neq T(L_c).$$

Therefore for all  $c \in C$  the distribution  $P_c$  has a density w.r.t.

$$(A.10) \quad \lambda = \sum_{\underline{c} \in C} \lambda_{\underline{c}} \quad :$$

$$(A.11) \quad \frac{dP_c}{d\lambda} = |D(c)|^{-1} I_{T(c)}(X) w_c(X).$$

This follows from  $\lambda_{\underline{c}} = \lambda_c \quad (\underline{c} \in D(c))$ , (A.7), (A.8), (A.9) leading to

$$(A.12) \quad \begin{aligned} & \int_H I_{T(c)}(X) w_c(X) d\lambda(X) = \\ & \sum_{\underline{c} \in D(c)} \int_H I_{H \cap T(c)}(X) w_c(X) d\lambda_{\underline{c}}(X) = |D(c)| \int_H w_c(X) d\lambda_c(X). \end{aligned}$$

Bayes formula together with (2.15) and (A.11) proves the theorem.

### 3. Proof of theorem 2.3:

The prior (2.4) and (A.3) give for  $P^X$ -almost all  $X \in T(c)$  the marginal density

$$(A.13) \quad p(\Lambda|c, X) \propto |\Lambda|^{(d_c+r-k)/2} e[\Lambda S_c]$$

under  $P_{c,X}$ . The conditional distribution of  $\begin{pmatrix} A \\ X \end{pmatrix}$  under the conditions  $U \in M_c$  and  $\Lambda$  is the normal  $N(G_0, \Lambda^{-1} \otimes \Psi_c)$ , where

$$(A.14) \quad G_0 = \begin{pmatrix} A_0 \\ Z A_0 \end{pmatrix}, \quad \Psi_c = \begin{pmatrix} \Gamma_0^{-1} & \Gamma_0^{-1} Z^t \\ Z \Gamma_0^{-1} & \Omega_c \end{pmatrix}.$$

The conditional density of  $A$  under the condition of a fixed  $X$  is known from Lemma 1 and yields together with (A.13) the density (2.31).

### 4. Proof of theorem 2.4:

The posterior mean (2.44) follows from (2.8) together with (2.17) and (2.31) because of

$$(A.15) \quad E_{c,X}(N(z_i A, \Lambda^{-1}) | \Lambda) = N(z_i \tilde{A}_c, (h_{ic} + 1)\Lambda^{-1})$$

leading with (A.6) to the density (2.45) of  $E_{c,X} N(z_i A, \Lambda^{-1})$ ; and

$$(A.16) \quad E_{c,X} \delta_{\tilde{Y}_j} = E_{c,X} N(X_i(j, c), h_{ijc}^2 \Lambda^{-1}) \quad , \quad j \in J_{ic}$$

leading to the density (2.48) of the distribution  $P_{ic}$ . The posterior means (2.38) follow from (2.5) together with (2.31), (A.1) and (A.2) and

$$\begin{aligned} (A.17) \quad \tilde{\Sigma}_c &= \omega E_{c,X} \Lambda^{-1/2} C_{a^{-1}\alpha_Y} \Lambda^{-1/2} = \\ &= \omega E_{c,X} \left[ \kappa \Lambda^{-1/2} I \Lambda^{-1/2} + \bar{\kappa} n^{-1} \sum_i (X_i - z_i A)^t (X_i - z_i A) - \right. \\ &\quad \left. - \bar{\kappa}^2 (\bar{x} - \bar{z} A)^t (\bar{x} - \bar{z} A) \right] = \\ &= \omega \left[ \kappa \widetilde{\Lambda_c^{-1}} + \bar{\kappa} E_{c,X} \left\{ R_c + n^{-1} \sum_i (z_i \Delta_c z_i^t) \Lambda^{-1} - \right. \right. \\ &\quad \left. \left. - \bar{\kappa} \left[ (\bar{x} - \bar{z} \tilde{A}_c)^t (\bar{x} - \bar{z} \tilde{A}_c) + (\bar{z} \Delta_c \bar{z}^t) \Lambda^{-1} \right] \right\} \right] = (2.39). \end{aligned}$$

### 5. Proof of theorem 4.1:

The proposition (4.9) follows from (2.31) and (2.36) by a cumbersome algebraic calculation using  $v_i = w_i^{-1}$  and the matrices

$$(A.18) \quad \Omega_{c_*} = \text{Diag}[v_1 J_{n_1}, v_2 J_{n_2}] + I \quad , \quad J_m = \mathbb{1}_m \mathbb{1}_m^t ,$$

$$(A.19) \quad \Omega_{c(n_1, n_2)} = \begin{pmatrix} I_{m_1} + v_1 J_{m_1} & v_1 \mathbb{1} & 0 & 0 \\ v_1 \mathbb{1}^t & 1 + v_1 & 0 & 1 \\ 0 & 0 & I_{m_2} + v_2 J_{m_2} & v_2 \mathbb{1} \\ 0 & 1 & v_2 \mathbb{1}^t & 1 + v_2 \end{pmatrix} .$$

Their inverses are

$$(A.20) \quad \Omega_{c_*}^{-1} = \text{Diag}[I_{n_1} - (n_1 + w_1)^{-1} J_{n_1} \quad , \quad I_{n_2} - (n_2 + w_2)^{-1} J_{n_2}]$$

and

$$(A.21) \quad \Omega_{c(n_1, n_2)}^{-1} = d \begin{pmatrix} d^{-1} I_{m_1} - J_{m_1} & -\mu_2 \mathbb{1} & -J & \lambda_2 \mathbb{1} \\ -\mu_2 \mathbb{1} & \mu_2 \lambda_1 & \lambda_1 \mathbb{1}^t & -\lambda_1 \lambda_2 \\ -J & \lambda_1 \mathbb{1} & d^{-1} I_{m_2} - J & -\mu_1 \mathbb{1} \\ \lambda_2 \mathbb{1}^t & -\lambda_1 \lambda_2 & -\mu_1 \mathbb{1}^t & \mu_1 \lambda_2 \end{pmatrix} ,$$

where

$$(A.22) \quad \mu_i := \lambda_i + 1 = n_i + w_i \quad (i = 1, 2).$$

Their determinants are

$$(A.23) \quad |\Omega_{c_*}| = (v_1 n_1 + 1)(v_2 n_2 + 1) \quad , \quad |\Omega_{c(i, j)}| = (w_1 w_2 d)^{-1}.$$

We remark, that the matrices  $\Omega_{c(i, j)}$  are obtained from  $\Omega_{c(n_1, n_2)}$  by row and column permutations.

## 6. Proof of theorem 4.2:

Following the definition of selfinformative limits given in section 3 we take  $r$  replications of the two-sample model, that is, we replace in (1.1), (4.1)  $n_j$  by  $rn_j$ ,  $Z$  by  $\mathbb{1}_r \otimes Z$ , use the observation matrix  $\mathbb{1}_r \otimes X$  and take the limit of the corresponding Bayes estimates (4.9) for  $r \rightarrow \infty$ , assuming  $X \in T$ .

For  $r \rightarrow \infty$  we obtain

$$(A.24) \quad \tilde{b}_{c_*m}^t \rightarrow \bar{x}_j \quad (m = 1, 2)$$

and for  $(i, j) \in V$

$$(A.25) \quad \tilde{b}_{c(i,j)1}^t \rightarrow h_{ij1} = \bar{x} + n_2^{-1}n(X_i - X_j)$$

$$(A.26) \quad \tilde{b}_{c(i,j)2}^t \rightarrow h_{ij2} = \bar{x} + n_1^{-1}n(X_j - X_i)$$

Moreover we have:

$$(A.27) \quad r^{-1}Q_* \rightarrow \bar{Q} = \sum_i X_i^t X_i - n_1 \bar{x}_1^t \bar{x}_1 - n_2 \bar{x}_2^t \bar{x}_2 \quad ,$$

$$(A.28) \quad r^{-1}Q_{i,j} \rightarrow \bar{Q}_{ij} = \bar{Q} + n_1(\bar{x}_1 - h_{ij1})^t(\bar{x}_1 - h_{ij1}) + n_2(\bar{x}_2 - h_{ij2})^t(\bar{x}_2 - h_{ij2}).$$

We see for  $(i, j) \in V$ , that  $\bar{Q}_{ij} - \bar{Q}$  is positive semidefinite and that

$$(A.29) \quad \bar{Q}_{ij} = \bar{Q} \text{ if and only if } h_{ijm} = \bar{x}_m \quad (m = 1, 2).$$

With

$$(A.30) \quad (i, l) \in V_1 = \{(i, l) \in V \mid h_{ilm} = x_m \quad (m = 1, 2)\}$$

we obtain for  $r \rightarrow \infty$  the limit

$$(A.31) \quad s_{c_*}^{-1} s_{c(i,j)} = q_{ij}(r)^{p/2} \longrightarrow q = (n^{-1} n_1 n_2)^{p/2} \quad ,$$

where

$$(A.32) \quad q_{ij}(r) = \prod_{m=1}^2 (n_m w_m^{-1} + r^{-1}) w_m (n + r^{-1} [w_1 + w_2 - 1])^{-1}.$$

With  $(i, j) \in V_2 = V - V_1$  it holds that  $|\overline{Q}| < |\overline{Q}_{i,j}|$  and therefore that

$$(A.33) \quad s_{c_*}^{-1} s_{c(i,j)} = q_{ij}(r) \left[ |r^{-1} S_0 + r^{-1} Q_*| |r^{-1} S_0 + r^{-1} Q_{i,j}|^{-1} \right]^{(g+rn)/2} \rightarrow 0.$$

From (A.31) and (A.33) follows that

$$(A.34) \quad r_{c(i,j)} \rightarrow 0 \quad [(i, j) \in V_2]$$

$$(A.35) \quad r_{c_*} \rightarrow \varrho = (1 + |V_1|q)^{-1} \quad , \quad r_{c(i,j)} \rightarrow \varrho q \quad [(i, j) \in V_1]$$

Finally (A.28), (A.34), (A.35) yield

$$(A.36) \quad \tilde{b}_m^t = \sum_{c \in \{c_*\} \cup V_1 \cup V_2} r_c \tilde{b}_{cm}^t \rightarrow \bar{x}_m.$$