

MIT LIBRARIES



3 9080 02618 0916

BASEMENT

HB31
.M415
no.07-10
2006

DEWEY

B31
M415
0.07-
10

Massachusetts Institute of Technology
Department of Economics
Working Paper Series

**POSTERIOR INFERENCE IN CURVED EXPONENTIAL FAMILIES
UNDER INCREASING DIMENSIONS**

Alexandre Belloni
Victor Chernozhukov

Working Paper 07-10
March 1, 2007

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

This paper can be downloaded without charge from the
Social Science Research Network Paper Collection at
<http://ssrn.com/abstract=970624>

POSTERIOR INFERENCE IN CURVED EXPONENTIAL FAMILIES UNDER INCREASING DIMENSIONS

BY ALEXANDRE BELLONI*

IBM T.J. Watson Research Center and MIT Operations Research Center
AND

BY VICTOR CHERNOZHUKOV†

MIT Department of Economics and Operations Research Center

In this work we study the large sample properties of the posterior-based inference in the curved exponential family under increasing dimension. The curved structure arises from the imposition of various restrictions, such as moment restrictions, on the model, and plays a fundamental role in various branches of data analysis. We establish conditions under which the posterior distribution is approximately normal, which in turn implies various good properties of estimation and inference procedures based on the posterior. We also discuss the multinomial model with moment restrictions, that arises in a variety of econometric applications. In our analysis, both the parameter dimension and the number of moments are increasing with the sample size.

1. Introduction. The main motivation for this paper is to obtain large sample results for posterior inference in the curved exponential family under increasing dimension. Recall that in the exponential family, the log of a density is linear in parameters $\theta \in \Theta$; in the curved exponential family, these parameters θ are restricted to lie on a curve $\eta \mapsto \theta(\eta)$ parameterized by a lower dimensional parameter $\eta \in \Psi$. There are many classical examples of densities that fall in the curved exponential family; see for example Efron [8], Lehmann and Casella [14], and Bandorff-Nielsen [1]. Curved exponential densities have also been extensively used in applications [8, 13]. An example of the condition that puts a curved structure onto an exponential family is a moment restriction of the type:

$$\int m(x, \alpha) f(x, \theta) dx = 0,$$

*Research support from a National Science Foundation grant is gratefully acknowledged. IBM Herman Goldstein Fellowship is also gratefully acknowledged.

†Research support from a National Science Foundation grant is gratefully acknowledged. Sloan Foundation Research Fellowship and Castle Krob Chair are also gratefully acknowledged.

that restricts θ to lie on a curve that can be parameterized as $\{\theta(\eta), \eta \in \Psi\}$, where component $\eta = (\alpha, \beta)$ contains α and other parameters β that are sufficient to parameterize all parameters $\theta \in \Theta$ that solve the above equation for some α . In econometric applications, often moment restrictions represent Euler equations that result from the data x being an outcome of an optimization by rational decision-makers; see e.g. Hansen and Singleton [9], Chamberlain [3], Imbens [11], and Donald, Imbens and Newey [5]. Thus, the curved exponential framework is a fundamental complement to the exponential framework, at least in certain fields of data analysis.

Despite the importance of applications to high-dimensional data, theoretical properties of the curved exponential family are not as well understood as the corresponding properties of the exponential family. In this paper, we contribute to the theoretical analysis of the posterior inference in curved exponential families of high dimension. We provide sufficient conditions under which consistency and asymptotic normality of the posterior is achieved when both the dimension of the parameter space and the sample size are increasing i.e. “large.” We allow for weak conditions on the priors, and expressly allow for improper priors in particular. We also study the convergence of moments and the precisions with which we can estimate them. We then apply these results to the multinomial model with moment restrictions, where both the parameter dimension and the number of moments are increasing with the sample size.

The present analysis of the posterior inference in the curved exponential family builds upon the previous work of Ghosal [12] who studied posterior inference in the exponential family under increasing dimension. Under sufficient growth restrictions on the dimension of the model, Ghosal showed that the posterior distributions concentrate in neighborhoods of the true parameter and can be approximated by an appropriate normal distribution. Ghosal’s analysis extended in a fundamental way the classical results of Portnoy [17] for maximum likelihood methods for the exponential family with increasing dimensions.

In addition to a detailed treatment of the curved exponential family, we also establish some useful results for exponential families. In fact, we begin our analysis by first revisiting Ghosal’s increasing dimension for the exponential family. We present several results that complement Ghosal’s results in several ways: First, we amend the conditions on priors to allow for a larger set of priors, for example, improper priors; second, we use concentration inequalities for log-concave densities to sharpen the conditions under which the normal approximations apply; and third, we show that the approximation of α -th order moments of the posterior by the corresponding moments

of the normal density becomes exponentially difficult in the order α .

The rest of the paper is organized as follows. In Section 2 we formally define the framework, assumptions, and develop results for the exponential family. In Section 3, the main section, we develop the results for the curved exponential family. In Section 4 we apply our results to the multinomial model with moment restrictions. Appendices C, D, and B collect proofs of the main results and technical lemmas.

2. Exponential Family Revisited. Assume that we have a triangular array of random samples

$$\begin{array}{ccccccc} X_1^{(1)} & & & & & & \\ X_1^{(2)} & X_2^{(2)} & & & & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \\ X_1^{(n)} & X_2^{(n)} & \dots & X_n^{(n)} & & & \end{array}.$$

Assume further that each $\{X_i^{(n)}\}_{i=1}^n$ are independent $d^{(n)}$ -dimensional vectors draw from a $d^{(n)}$ -dimensional standard exponential family whose density is defined by

$$(2.1) \quad f(x; \theta^{(n)}) = \exp\left(\langle x, \theta^{(n)} \rangle - \psi^{(n)}(\theta^{(n)})\right),$$

where $\theta^{(n)} \in \Theta^{(n)}$ an open convex set of $\mathbb{R}^{d^{(n)}}$ and $\psi^{(n)}$ is the associate normalizing function. Let $\theta_0^{(n)} \in \Theta^{(n)}$ denote the (sequence of) true parameter which is assumed to be bounded away from the boundary of $\Theta^{(n)}$ (uniformly in n). For notational convenience we will suppress the superscript (n) but is understood that the associate objects are changing with n .

Under this framework, the posterior density of θ given the observed data $\{X_i\}_{i=1}^n$ is defined as

$$(2.2) \quad \pi_n(\theta) \sim \pi(\theta) \prod_{i=1}^n f(X_i; \theta) = \pi(\theta) \exp\left(\left\langle \sum_{i=1}^n X_i, \theta \right\rangle - n\psi(\theta)\right),$$

where $\pi (= \pi^{(n)})$ denotes a prior distribution on Θ . As expected, we will need to impose some regularity conditions on the prior π . These conditions differs from the ones imposed in [12]. Although the same Lipschitz condition is required, we require only a relative lower bound on the value of the prior on the true parameter instead of an absolute bound (see Theorem 1). Finally, our conditions allow for improper priors in opposition to [12]. In fact, the uninformative prior trivially satisfies our assumptions.

Our results are stated in terms of a re-centered gaussian distribution in the local parameter space. Let $\mu = \psi(\theta_0)$ and $F = \psi''(\theta_0)$ be the mean and covariance matrix associated with the random variables $\{X_i\}$, and let $J = F^{1/2}$ be its square root (i.e., $JJ^T = F$). The re-centering is defined as $\Delta_n := \sqrt{n}J^{-1}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mu\right)$; it follows that $E[\Delta_n] = 0$, and $E[\Delta_n\Delta_n^T]$ is the identity matrix of appropriate (increasing) dimension d . Moreover, the posterior in the local parameter space is defined for $u \in \mathcal{U} = \sqrt{n}J(\Theta - \theta_0)$ as

$$(2.3) \quad \pi^*(u) = \frac{\pi(\theta_0 + n^{-1/2}J^{-1}u) \prod_{i=1}^n f(X_i; \theta_0 + n^{-1/2}J^{-1}u)}{\int_{\mathcal{U}} \pi(\theta_0 + n^{-1/2}J^{-1}u) \prod_{i=1}^n f(X_i; \theta_0 + n^{-1/2}J^{-1}u) du}.$$

In the same lines of Portnoy [17] and Ghosal [12], conditions on the growth rates of the third and fourth moments are required. More precisely, the following quantities play an important role in the analysis:

$$(2.4) \quad B_{1n}(c) = \sup \left\{ E_{\theta} \left[\langle a, V \rangle^3 \right] : a \in S^{d-1}, \|J(\theta - \theta_0)\|^2 \leq \frac{cd}{n} \right\},$$

$$(2.5) \quad B_{2n}(c) = \sup \left\{ E_{\theta} \left[\langle a, V \rangle^4 \right] : a \in S^{d-1}, \|J(\theta - \theta_0)\|^2 \leq \frac{cd}{n} \right\},$$

where V is a random variable distributed as $J^{-1}(U - E_{\theta}[U])$ and U has density $f(\cdot; \theta)$ as defined in (2.1). Moreover, the following combination of (2.4) and (2.5) is relevant

$$(2.6) \quad \lambda_n(c) := \frac{1}{6} \left(\sqrt{\frac{cd}{n}} B_{1n}(0) + \frac{cd}{n} B_{2n}(c) \right),$$

where we note that $\lambda_n(c)$ is different (in fact smaller) than the one defined in [12]. This quantity is used to bound deviations from normality of the posterior in a neighborhood of the true parameter.

Next we state the main results of this work.

Theorem 1 *For any constant $c > 0$ suppose that:*

- (i) $B_{1n}(c)\sqrt{d/n} \rightarrow 0$;
- (ii) $\lambda_n(c)d \rightarrow 0$;
- (iii) $\|F^{-1}\|d/n \rightarrow 0$;
- (iv) *the prior π density satisfies: $\sup_{\theta} \ln \frac{\pi(\theta)}{\pi(\theta_0)} \leq O(d)$, and*

$$|\ln \pi(\theta) - \ln \pi(\theta_0)| \leq K_n(c) \|\theta - \theta_0\|$$

for any θ such that $\|\theta - \theta_0\| \leq \sqrt{\|F^{-1}\|cd/n}$. We require $K_n(c)\sqrt{\|F^{-1}\|cd/n} \rightarrow 0$.

Then we have asymptotic normality of the posterior density function, that is

$$\int |\pi_n^*(u) - \phi_d(u; \Delta_n, I_d)| du \rightarrow_p 0.$$

As mentioned earlier, Theorem 1 has different assumptions on the prior that Theorem 3 of [12] has. On the other hand, it does not require additional technical assumptions used in [12], as discussed in Section B, and the growth condition of d with relative to the sample size n is improved by $\ln d$ factors.

In some applications it might be desired to have stronger convergence properties than simply asymptotic normality. The following theorem provides sufficient conditions for the α -moment convergence.

Theorem 2 For some sequence of $\{\alpha\}$ and $d \rightarrow \infty$, let

$$M_{d,\alpha} := (d + \alpha) \left(1 + \frac{\alpha \ln(d + \alpha)}{d + \alpha} \right).$$

Suppose that the following strengthening of assumptions (ii) and (iv) hold for any fixed \bar{c} :

$$(ii') \quad \lambda_n (\bar{c} M_{d,\alpha}/d) \left[\bar{c} M_{d,\alpha} \right]^{1+\alpha/2} \rightarrow 0;$$

$$(iv') \quad K_n (\bar{c} M_{d,\alpha}/d) \sqrt{\frac{\|F^{-1}\| [\bar{c} M_{d,\alpha}]^{1+\alpha/2}}{n}} \rightarrow 0.$$

Then we have

$$(2.7) \quad \int \|u\|^\alpha |\pi_n^*(u) - \phi_d(u; \Delta_n, I_d)| du \rightarrow_p 0.$$

We emphasize that Theorem 2 allows for α and d to grow as the sample size increases. Our conditions highlight the polynomial trade off between n and d but an exponential trade off between n and α . This suggests that the estimation of higher moments in increasing dimensions applications could be very delicate. Conditions (ii') and (iv') simplify significantly if $\alpha = o(d)$, in such case we have $M_{d,\alpha} \sim d$.

As an illustrative example consider the multinomial distribution application analyzed by Ghosal in [12]. Let $\mathcal{X} = \{x^0, x^1, \dots, x^d\}$ be the known finite support of a multinomial random variable X where d is allowed to grow with sample size n . For each i denote by p_i the probability of the event $\{X = x^i\}$ which is assumed to satisfy $\max_i 1/p_i = O(d)$. The parameter space is given by $\theta = (\theta_1, \dots, \theta_d)$ where $\theta_i = \log(p_i/(1 - \sum_{j=1}^d p_j))$ (under the assumption on the p_i 's the true value of θ_j 's is bounded). The Fisher

information matrix is given by $F = P - pp'$ where $P = \text{diag}(p)$. Using a rank-one update formula, we have

$$(2.8) \quad F^{-1} = P^{-1} + \frac{P^{-1}pp'P^{-1}}{1 - p'P^{-1}p} = P^{-1} + \frac{ee'}{1 - p_0}.$$

Therefore we have $\|F^{-1}\| \leq \text{trace}(F^{-1}) = \sum_{i=1}^d \frac{1}{p_i} + \frac{d}{1-p_0} \leq O(d^2)$. It is also possible to derive an expression for $J = F^{1/2}$ and its inverse

$$J = P^{1/2} - \frac{pp'P^{-1/2}}{1 + \sqrt{1 - p'P^{-1}p}} \quad \text{and} \quad J^{-1/2} = P^{-1/2} + \frac{P^{-1}pp'P^{-1/2}}{1 - p'P^{-1}p + \sqrt{1 - p'P^{-1}p}}$$

In order to bound λ_n we need to bound the third and fourth moments of a random variable which define B_{1n} and B_{2n} . Let $a \in S^{n-1}$, and q be distributed as $f(\cdot; \theta)$. We have that

$$\langle a, J^{-1}(q - p) \rangle = \sum_{i=1}^d a_i p_i^{-1/2} (q_i - p_i) + \frac{e'a}{p_0 + \sqrt{p_0}} \sum_{i=1}^d p_i^{1/2} (q_i - p_i).$$

Under the assumption on the p_i it can be shown that $B_{1n}(c) = O(d^{3/2})$ and $B_{2n}(c) = O(p^2)$.

The relations above were derived by Ghosal in [12], where the growth condition that $d^6(\ln d)/n \rightarrow 0$ was imposed to obtain the asymptotic normality results (the case of $\alpha = 0$). We relax this growth requirement by combining Ghosal's approach with our analysis and an uninformative (improper) prior. In this case we have $K_n(c) = 0$ and our definition of λ_n remove the logarithmic factors. Therefore, Theorem 1 leads to a weaker growth condition in that it only requires that $d^4/n \rightarrow 0$. Moreover, the results of Theorem 2.4 of [12] now follow under the weaker growth condition that $d^5/n \rightarrow 0$, replacing the previous growth condition that $d^6(\log d)/n \rightarrow 0$. For higher moment estimation ($\alpha > 0$), the conditions of Theorem 2 are satisfied with the condition that $d^{4+\alpha+\delta}/n \rightarrow 0$ for any strictly positive value of δ .

Suppose next that we are interested to allow α grow with the sample size as well. If d is growing in a polynomial rate with respect to n , our results do not allow for $\alpha = O(\ln n)$. Some limitation along these lines should be expected since there is an exponential trade off between α and n . However, it is definitely possible to let both the dimension and α to grow with the sample size if, for instance, we are willing to accept the condition that $\alpha = O(\sqrt{\ln n})$. Such slow growth conditions illustrate the potential limitations for the practical estimation of higher order moments.

3. Curved Exponential Family. Next we consider the case of a curved exponential family. Being a generalization of the canonical exponential family, its analysis has many similarities with the previous setup.

Let X_1, X_2, \dots, X_n be iid observations from a d -dimensional curved exponential family whose density function is given by

$$f(x; \theta) = \exp(\langle x, \theta(\eta) \rangle - \psi_n(\theta(\eta))),$$

where $\eta \in \Psi \subset \mathbb{R}^{d_1}$, $\theta : \Psi \rightarrow \Theta$, an open subset of \mathbb{R}^d , and $d \rightarrow \infty$ as $n \rightarrow \infty$ as before. In this section we assume that $J = I_d$ for notational convenience.

The parameter of interest is η , whose true value η_0 lies in the interior of the set $\Psi \subset \mathbb{R}^{d_1}$. The true value of θ induced by η_0 is given by $\theta_0 = \theta(\eta_0)$. The mapping $\eta \mapsto \theta(\eta)$ takes values from \mathbb{R}^{d_1} to \mathbb{R}^d where $c \cdot d \leq d_1 \leq d$, for some $c > 0$. Moreover, assume that η_0 is the unique solution to the system $\theta(\eta) = \theta_0$.

Thus, the parameter θ corresponds to a high-dimensional linear parametrization of the log-density, and η describes the lower-dimensional parametrization of the density of interest. We require the following regularity conditions on the mapping $\theta(\cdot)$.

Assumption A. For every κ , and uniformly in $\gamma \in B(0, \kappa\sqrt{d})$, there exists a linear operator $G : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^d$ such that $G'G$ has eigenvalues bounded from above and away from zero, and for every n

$$(3.9) \quad \sqrt{n}(\theta(\eta_0 + \gamma/\sqrt{n}) - \theta(\eta_0)) = r_{1n} + (I + R_{2n})G\gamma,$$

where $\|r_{1n}\| \leq \delta_{1n}$ and $\|R_{2n}\| \leq \delta_{2n}$. Moreover, those coefficients are such that

$$(3.10) \quad \delta_{1n}d^{1/2} \rightarrow 0 \quad \text{and} \quad \delta_{2n}d \rightarrow 0.$$

Assumption B. There exist a strictly positive constants ε_0 such that for every $\eta \in \Psi$ (uniformly on n) we have

$$(3.11) \quad \|\theta(\eta) - \theta(\eta_0)\| \geq \varepsilon_0 \|\eta - \eta_0\|.$$

Thus the mapping $\eta \mapsto \theta(\eta)$ is allowed to be nonlinear and discontinuous. For example, the additional condition of $\delta_{1n} = 0$ implies the continuity of the mapping in a neighborhood of η_0 . More generally, condition (3.10) does impose that the map admits an approximate linearization in the neighborhood of η_0 whose quality is controlled by the errors δ_{1n} and δ_{2n} . An example of a kind of map allowed in this framework is given in the figure.

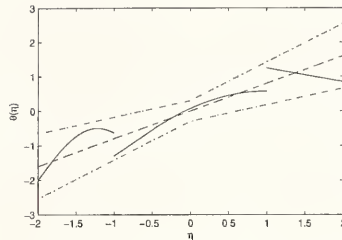


FIG 1. This figure illustrates the mapping $\theta(\cdot)$. The (discontinuous) solid line is the mapping while the dash line represents the linear map induced by G . The dash-dot line represents the deviation band controlled by r_{1n} and R_{2n} .

Again, given a prior π on Θ , the posterior of η given the data is denoted by

$$\pi_n(\eta) \propto \pi(\theta(\eta)) \cdot \prod_{i=1}^n f(x_i; \eta) = \pi(\theta(\eta)) \cdot \exp \{n \langle \bar{X}, \theta(\eta) \rangle - n\psi(\theta(\eta))\}$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

Under this framework, we also define the local parameter space to describe contiguous deviations from the true parameter as

$$\gamma = \sqrt{n}(\eta - \eta_0), \text{ and let } s = (G'G)^{-1}G'\sqrt{n}(\bar{X} - \mu)$$

(once more) be a first order approximation to the normalized maximum likelihood/extremum estimate. Again, similar bounds hold for s : $E[s] = 0$, $E[ss^T] = (G'G)^{-1}$, and $\|s\| = O_p(\sqrt{d})$. The posterior density of γ over Γ , where $\Gamma = \sqrt{n}(\Psi - \eta_0)$, is $\pi_n^*(\gamma) = \frac{\ell(\gamma)}{\int_{\Gamma} \ell(\gamma) d\gamma}$, where

(3.12)

$$\begin{aligned} \ell(\gamma) &= \exp \left(n \langle \bar{X}, \theta(\eta_0 + n^{-1/2}\gamma) - \theta(\eta_0) \rangle - n \left[\psi(\theta(\eta_0 + n^{-1/2}\gamma)) - \psi(\theta(\eta_0)) \right] \right) \\ &\quad \times \pi \left(\theta \left(\eta_0 + n^{-1/2}\gamma \right) \right). \end{aligned}$$

By construction we have

$$\frac{\ell(\gamma)}{\pi \left(\theta \left(\eta_0 + n^{-1/2}\gamma \right) \right)} = Z_n \left(n^{1/2} \left[\theta \left(\eta_0 + n^{-1/2}\gamma \right) - \theta(\eta_0) \right] \right) = Z_n(u_\gamma)$$

where $u_\gamma \in \mathcal{U} \subset \mathbb{R}^d$.

Next we first show that tails have small mass outside a \sqrt{d} -neighborhood in Γ . We also need an additional condition on a_n as defined in (B.18) and repeated here for the reader's convenience

$$a_n = \sup\{c : \lambda_n(c) \leq 1/16\}.$$

Therefore, using Lemma 2, in a neighborhood of size $\sqrt{a_n d}$ we can still bound Z_n by above with a proper gaussian. In the next lemma it is required that $\log d = o(a_n)$ which is a substantially weaker condition than the one used in [12] for establishing asymptotic normality for the posterior of (regular) exponential densities, $\lambda_n(c \log d)d = o(1)$.

Lemma 1 *Assume that (i), (ii), (iii), and (iv) hold. In addition, suppose that $\log d = o(a_n)$. Then, for some constant \bar{k} independent of d and d_1 , we have*

$$\int_{\Gamma \setminus B(0, \bar{k}\sqrt{d})} \pi\left(\theta(\eta_0) + n^{-1/2}u_\gamma\right) Z_n(u_\gamma) d\gamma \leq o\left(\int_{\Gamma} \pi\left(\theta(\eta_0) + n^{-1/2}u_\gamma\right) Z_n(u_\gamma) d\gamma\right)$$

Comment 3.1 *The only assumption made on d_1 in the previous lemma was that $d_1 \leq d$. If $d_1 \log d = o(d)$ the proof simplifies significantly (there is no need to define region (II)).*

Next we address the consistency question for the maximum likelihood estimator associated with the curved exponential family.

Theorem 3 *In addition to Assumptions A and B, suppose that $a_n \rightarrow \infty$, and (iv) hold. Then the maximum likelihood estimator $\hat{\eta}$ satisfies*

$$\|\hat{\eta} - \eta_0\| = O(\sqrt{d/n}).$$

Two remarks regarding Theorem 3 are worth mention. First, a sufficient condition for $a_n \rightarrow \infty$ is simply $\lambda_n(c) \rightarrow 0$, stronger than the condition $\sqrt{d/n}B_{1n}(c)$ needed for consistency for the exponential case obtained by Ghosal in [12]. Second, our consistency result relies on the dimension of the larger model d .

Finally, we can state the asymptotic normality result for the curved exponential family.

Theorem 4 *Suppose that Assumptions A, B, (i), (ii), (iii), and (iv) hold. In addition, suppose that $\log d = o(a_n)$. Then, asymptotic normality for the posterior density associated with the curved exponential family holds,*

$$\int |\pi_n^*(\gamma) - \phi_{d_1}(\gamma; s, (G'G)^{-1})| d\gamma.$$

4. Multinomial Model with Moment Restrictions. In this section we provide a high-level discussion of the multinomial model with moment restrictions. Let $\mathcal{X} = \{x^0, x^1, x^2, \dots, x^d\}$ be the known finite support of a multinomial random variable X which was described in Section 2. Conditions (i) – (iv) were verified in the same section.

As discussed in the introduction, it is of interest to incorporate moment restrictions into this model, see Imbens [11] for a discussion. This will lead to a curved exponential model as studied in Section 3.

The parameter of interest is $\eta \in \Psi \subset \mathbb{R}^{d_1}$ a compact set. Consider a (twice continuously differentiable) vector-valued moment function $m : \mathcal{X} \times \Psi \rightarrow \mathbb{R}^M$ such that

$$\mathbb{E}[m(X, \eta)] = 0 \quad \text{for a unique } \eta_0 \in \Psi.$$

The case of interest consists of the cardinality d of the support \mathcal{X} being larger than the number of moment conditions M which in turn is larger than the dimension d_1 of the parameter of interest η . The log-likelihood function associated with this model

$$(4.13) \quad \begin{aligned} l(q, \eta) &= \sum_{i=1}^n \sum_{j=0}^d I\{X_i = x_j\} \ln q_j \\ &\text{for some } \theta \text{ and } \eta \text{ such that } \sum_{j=0}^d q_j m(x_j, \eta) = 0, \sum_{j=0}^d q_j = 1. \end{aligned}$$

and $l(q, \eta) = -\infty$ if violates any of the moments conditions. This log-likelihood function induces the mapping $q : \Psi \rightarrow \Delta^{d-1}$ formally defined as

$$(4.14) \quad \begin{aligned} q(\eta) &= \arg \max_q l(q, \eta) \\ &\sum_{j=0}^d q_j m(x_j, \eta) = 0, \sum_{j=1}^d q_j = 1, q \geq 0. \end{aligned}$$

As discussed in Section 2 the function $\theta_j(\eta) = \log(q_j(\eta)/q_0(\eta))$ (for $j = 1, \dots, d$) is the natural $\theta(\cdot) : \Psi \rightarrow \Theta$ mapping. Assuming that the matrix $\mathbb{E}[m(X, \eta)m(X, \eta)']$ is uniformly positive definite over η , Qin and Lawless [18] use the inverse function theorem to show that θ is a twice continuous differentiable mapping of η in a neighborhood of η_0 . In particular this implies that Assumption A holds with $\delta_{2n} = 0$ and $\delta_{1n} = O(d_1^2(d/n))$. It suffices to have $d^{4.5}/n \rightarrow 0$.

In order to verify Assumption B, we use that the parameter η belongs in a compact set Ψ , and assume that the mapping is injective (over a set that contains Ψ in its interior). We refer to Newey and McFadden [16] for a discussion of primitive assumptions for identification with moment restrictions.

APPENDIX A: NOTATION

For $a, b \in \mathbb{R}^d$, their (Euclidean) inner product is denoted by $\langle a, b \rangle$, and $\|a\| = \sqrt{\langle a, a \rangle}$. The unit sphere in \mathbb{R}^d is denoted by $S^{d-1} = \{v \in \mathbb{R}^d : \|v\| = 1\}$. For a linear operator A , the operator norm is denoted by $\|A\| = \sup\{\|Aa\| : \|a\| = 1\}$. Let $\phi_d(\cdot; \mu; V)$ denote the d -dimensional gaussian density function with mean μ and covariance matrix V .

APPENDIX B: TECHNICAL RESULTS

In this section we prove the technical lemmas needed to prove our main result in the following section. Our exposition follows the work of Ghosal [12]. For the sake of completeness we include Proposition 1, which can be found in Portnoy [17], and a specialized version of Lemma 1 of Ghosal [12]. All the remaining proofs use different techniques and weaker assumptions which leads to a sharper analysis. In particular, we no longer require the prior to be proper, no bounds on the growth of $\det(\psi''(\theta_0))$ are imposed, and $\ln n$ and $\ln d$ do not need to be of the same order.

As mentioned earlier we follow the notation in Ghosal [12], for $u \in \mathcal{U}$ let

$$(B.15) \quad \tilde{Z}_n(u) = \exp\left(\langle u, \Delta_n \rangle - \|u\|^2/2\right) \quad \text{and}$$

$$(B.16) \quad Z_n(u) = \exp\left(\frac{1}{\sqrt{n}} \left\langle \sum_{i=1}^n X_i, J^{-1}u \right\rangle - n \left[\psi\left(\theta_0 + n^{-1/2} J^{-1}u\right) - \psi(\theta_0) \right]\right),$$

otherwise (if $\theta_0 + n^{-1/2} J^{-1}u \notin \Theta$), let $Z_n(u) = \tilde{Z}_n(u) = 0$. The quantity (B.15) denotes the likelihood ratio associated with f as a function of u . In a parallel manner, (B.16) is associated with a standard gaussian density.

We start recalling a result on the Taylor expansion of ψ which is key to control deviations between $\tilde{Z}(u)$ and $Z(u)$.

Proposition 1 (Portnoy [17]) *Let ψ' and ψ'' denote respectively the gradient and the Hessian of ψ . For any $\theta, \theta_0 \in \Theta$, there exists $\tilde{\theta} = \lambda\theta + (1-\lambda)\theta_0$, for some $\lambda \in [0, 1]$, such that*

$$(B.17) \quad \begin{aligned} \psi(\theta) &= \psi(\theta_0) + \langle \psi'(\theta_0), \theta - \theta_0 \rangle + \frac{1}{2} \langle \theta - \theta_0, \psi''(\theta_0)(\theta - \theta_0) \rangle + \\ &\quad + \frac{1}{6} E_{\theta_0} \left[\langle \theta - \theta_0, U \rangle^3 \right] + \frac{1}{24} \left\{ E_{\tilde{\theta}} \left[\langle \theta - \theta_0, U \rangle^4 \right] - 3 \left(E_{\tilde{\theta}} \left[\langle \theta - \theta_0, U \rangle^2 \right] \right)^2 \right\} \end{aligned}$$

where $E_{\tilde{\theta}}[g(U)]$ denotes the expectation of $g(V - E[V])$ and $V \sim f(\cdot; \tilde{\theta})$.

Based on Proposition 1 we control the pointwise deviation between Z_n and \tilde{Z}_n in a neighborhood of zero (i.e., in a neighborhood of the true parameter).

Lemma 2 (Essentially in Ghosal [12] or Portnoy [17]) *For all u such that $\|u\| \leq \sqrt{cd}$, we have*

$$|\ln Z_n(u) - \ln \tilde{Z}_n(u)| \leq \lambda_n(c) \|u\|^2 \quad \text{and} \quad \ln Z_n(u) \leq \langle \Delta_n, u \rangle - \frac{1}{2} \|u\|^2 (1 - 2\lambda_n(c)).$$

Proof. Under our definitions, $(I) = |\ln \tilde{Z}_n(u) - \ln Z_n(u)| = n|\psi(\theta_0 + n^{-1/2}J^{-1}u) - \psi(\theta_0)|$. Using Proposition 1 we have that (I) is bounded above by

$$\begin{aligned} (I) &\leq n \left| \frac{1}{6} E \left[\left\langle \frac{u}{n^{1/2}}, V \right\rangle^3 \right] + \frac{1}{24} \left\{ E_{\tilde{\theta}} \left[\left\langle \frac{u}{n^{1/2}}, V \right\rangle^4 \right] - 3 \left(E_{\tilde{\theta}} \left[\left\langle \frac{u}{n^{1/2}}, V \right\rangle^2 \right] \right)^2 \right\} \right| \\ &\leq \frac{1}{6} (n^{-1/2} \|u\|^3 B_{1n}(0) + n^{-1} \|u\|^4 B_{2n}(c)) \leq \lambda_n(c) \|u\|^2. \end{aligned}$$

The second inequality follows directly from the first result. ■

Next we show how to bound the integrated deviation between the quantities in (B.15) and (B.16) restricted to the neighborhood of zero.

Lemma 3 *For any $c > 0$ we have*

$$\left(\int \tilde{Z}_n(u) du \right)^{-1} \int_{\{u: \|u\| \leq \sqrt{cd}\}} |Z_n(u) - \tilde{Z}_n(u)| du \leq cd \lambda_n(c) e^{cd \lambda_n(c)}$$

Proof. Using $|e^x - e^y| \leq |x - y| \max\{e^x, e^y\}$ and Lemma 2 (since $\|u\| \leq \sqrt{cd}$) we have

$$\begin{aligned} |Z_n(u) - \tilde{Z}_n(u)| &\leq |\ln Z_n(u) - \ln \tilde{Z}_n(u)| \exp \left(\langle \Delta_n, u \rangle - \frac{1}{2} (1 - \lambda_n(c)) \|u\|^2 \right) \\ &\leq \lambda_n(c) \|u\|^2 \exp \left(\langle \Delta_n, u \rangle - \frac{1}{2} (1 - \lambda_n(c)) \|u\|^2 \right). \end{aligned}$$

By integrating over the set $H(\sqrt{cd}) = \{u : \|u\| \leq \sqrt{cd}\}$ we obtain

$$\begin{aligned} \int_{H(\sqrt{cd})} |Z_n(u) - \tilde{Z}_n(u)| du &\leq \int_{H(\sqrt{cd})} \lambda_n(c) \|u\|^2 \exp \left(\langle \Delta_n, u \rangle - \frac{1}{2} (1 - \lambda_n(c)) \|u\|^2 \right) \\ &\leq cd \lambda_n(c) \int_{H(\sqrt{cd})} \exp \left(\langle \Delta_n, u \rangle - \frac{1}{2} (1 - \lambda_n(c)) \|u\|^2 \right) \\ &\leq cd \lambda_n(c) e^{cd \lambda_n(c)} \int_{H(\sqrt{cd})} \exp \left(\langle \Delta_n, u \rangle - \frac{1}{2} \|u\|^2 \right) \\ &\leq cd \lambda_n(c) e^{cd \lambda_n(c)} \int \tilde{Z}_n(u) du. \end{aligned}$$

■

The next lemma controls the tail of Z_n relatively to \tilde{Z}_n . In order to achieve that it makes use of a concentration inequality for log-concave densities functions developed by Lovász and Vempala in [15]. The lemma is stated with a given bound on the norm of Δ_n which is allowed to grow with the dimension. Such bound on Δ_n can be easily obtained with probability arbitrary close to one by standard arguments.

Lemma 4 Suppose that $\|\Delta_n\|^2 < C_1 d$ and $\lambda_n(c) < 1/16$. Then for every $k \geq 1$ we have

$$\int_{\{u: \|u\| \geq k\sqrt{cd}\}} \pi(\theta_0 + n^{-1/2} J^{-1} u) Z_n(u) du \leq \left(\sup_u \pi(u) \right) \left(e^{cd\lambda_n(c)} \int \tilde{Z}_n(u) du \right) e^{-k\frac{c}{8}d}$$

where $c > 16 \max\{4C_1, 1/(1 - 2\lambda_n(c))\}$.

Proof. Define $H(a) := \{u : \|u\| \leq a\}$ and its complement by $H(a)^c$. Then we have

$$\int_{H(k\sqrt{cd})^c} \pi(\theta_0 + n^{-1/2} J^{-1} u) Z_n(u) du \leq \sup_{H(k\sqrt{cd})^c} \pi(\theta_0 + n^{-1/2} J^{-1} u) \int_{H(k\sqrt{cd})^c} Z_n(u) du.$$

Next note that $\Delta_n \in H(k\sqrt{cd})$. Moreover, for any $u \in H(k\sqrt{cd})^c$ we have some $\tilde{u} := k\sqrt{cd}u/\|u\|$ such that

$$\begin{aligned} \ln Z_n(u) &\leq \langle \Delta_n, \tilde{u} \rangle - \frac{1}{2} \|\tilde{u}\|^2 (1 - 2\lambda_n(c)) \\ &\leq k\sqrt{cd}\sqrt{C_1 d} - \frac{1}{2} k^2 cd (1 - 2\lambda_n(c)) \\ &\leq -kd \left(\frac{1 - 2\lambda_n(c)}{2} kc - \sqrt{cC_1} \right) \end{aligned}$$

Under our assumptions we have

$$\tilde{c} := \left(\frac{(1 - 2\lambda_n(c))}{2} kc - \sqrt{cC_1} \right) \geq kc \left(\frac{1}{2} - \lambda_n(c) - \frac{1}{8} \right) \geq 2$$

since $c > 16$ and assuming $\lambda_n(c) < 1/4$. Using Lemma 5.16 of [15] we have

$$\begin{aligned} \int_{H(k\sqrt{cd})^c} Z_n(u) du &\leq (e^{1-\tilde{c}})^{d-1} \int Z_n(u) du \\ &\leq 2(e^{1-\tilde{c}})^{d-1} \int_{H(\sqrt{cd})} Z_n(u) du \\ &\leq 2(e^{1-\tilde{c}})^{d-1} e^{cp\lambda_n(c)} \int_{H(\sqrt{cd})} \tilde{Z}_n(u) du \end{aligned}$$

where we used that $\int Z_n(u) du \leq 2 \int_{H(\sqrt{cd})} Z_n(u) du$ (note that k does not appear).

Since $c > 16$, we have $m := c \left((1 - 1/d)^{\frac{1}{5}} - \lambda_n(c) \right) - 1$ (since we expect $\lambda_n(c) \rightarrow 0$ and $1/d \rightarrow 0$, $m \rightarrow \frac{c}{5} - 1 \geq \frac{c}{8}$). ■

We note that the value of c in the previous lemma could depend on n as long as the condition is satisfied. In fact, we can have c as large as

$$(B.18) \quad a_n := \sup\{c : \lambda_n(c) < 1/16\}.$$

(B.18) characterizes a neighborhood of size $\sqrt{a_n d}$ on which the quantity $Z_n(\cdot)$ can still be bounded by a proper gaussian. Lemma 4 bounds the contribution outside this neighborhood. We close this section with a technical lemma for bounding the difference between the expectation of a function with respect to two probability densities.

Lemma 5 *Let f and g be two nonnegative integrable functions in \mathbb{R}^d , and define $I_f = \int f(u)du$ and $I_g = \int g(u)du$. Moreover, let h be a third positive function and $\Lambda \subset \mathbb{R}^d$ be a set such that $h_f = \int_{\Lambda} h(u)f(u)du < \infty$. Then*

$$\int_{\Lambda} h(u) \left| \frac{f(u)}{I_f} - \frac{g(u)}{I_g} \right| du \leq \frac{\max_{u \in \Lambda} h(u) + h_f/I_f}{I_g} \int_{\Lambda} |f(u) - g(u)| du.$$

Proof. Simply note that

$$\begin{aligned} \int_{\Lambda} h(u) \left| \frac{f(u)}{I_f} - \frac{g(u)}{I_g} \right| du &= \int_{\Lambda} h(u) \left| \frac{f(u)}{I_f} - \frac{f(u)}{I_g} + \frac{f(u)-g(u)}{I_g} \right| du \\ &\leq \left| \frac{1}{I_f} - \frac{1}{I_g} \right| \int_{\Lambda} h(u)f(u)du + \frac{1}{I_g} \int_{\Lambda} h(u)|f(u) - g(u)|du \\ &= \left| \frac{I_f - I_g}{I_f I_g} \right| h_f + \frac{1}{I_g} \int_{\Lambda} h(u)|f(u) - g(u)|du \\ &\leq \frac{\max_{u \in \Lambda} h(u) + h_f/I_f}{I_g} \int_{\Lambda} |f(u) - g(u)|du. \end{aligned}$$

■

APPENDIX C: PROOF OF THEOREMS 1 AND 2

Armed with Lemmas 2, 3, 4, and 5, we now show asymptotic normality and moments convergence results (respectively Theorems 1 and 2) under the appropriate growth conditions of the dimension of the parameter space with respect to the sample size.

It is easy to see that Theorem 1 follows from Theorem 2 with $\alpha = 0$, therefore its proof is omitted.

Proof of Theorem 2. Let $M_{d,\alpha} := (d + \alpha) \left(1 + \frac{\alpha \ln(d + \alpha)}{d + \alpha} \right)$. In the case that α is constant and d grows to infinity, this simplifies to a multiple of d . We will be using that $\sqrt{\bar{c}M_{d,\alpha}} \geq 4\|\Delta_n\|$ in the analysis (recall that $\|\Delta_n\| = O(\sqrt{d})$). We will divide the integral of (2.7) in two regions

$$\Lambda = \left\{ u \in \mathbb{R}^d : \|u\| \leq \sqrt{\bar{c}M_{d,\alpha}} \right\} \quad \text{and} \quad \Lambda^c,$$

where \bar{c} is a fixed constant. Thus we have

$$\begin{aligned} \text{(C.19)} \quad \int \|u\|^\alpha |\pi_n^*(u) - \phi_d(u; \Delta_n, I_d)| du &\leq \int_{\Lambda} \|u\|^\alpha |\pi_n^*(u) - \phi_d(u; \Delta_n, I_d)| du + \\ &\quad + \int_{\Lambda^c} \|u\|^\alpha |\pi_n^*(u) - \phi_d(u; \Delta_n, I_d)| du. \end{aligned}$$

To bound the first term, we will use Lemma 5 with $h(u) = \|u\|^\alpha$ and Λ as defined above. In this case, we have $h_f/I_f \leq \max_{u \in \Lambda} h(u) = \bar{c}^{\alpha/2} M_{d,\alpha}^{\alpha/2}$.

Therefore

$$\begin{aligned}
\int_{\Lambda} \|u\|^{\alpha} |\pi_n^*(u) - \phi_d(u; \Delta_n, I_d)| du &\leq \\
&\leq \frac{2\bar{c}^{\alpha/2} M_{d,\alpha}^{\alpha/2}}{\int \pi(\theta_0) \tilde{Z}_n(u) du} \int_{\Lambda} |\pi(\theta_0 + n^{-1/2} J^{-1} u) Z_n(u) - \pi(\theta_0) \tilde{Z}_n(u)| du \\
&\leq 2\bar{c}^{\alpha/2} M_{d,\alpha}^{\alpha/2} \sup_{u \in \Lambda} \left| \frac{\pi(\theta_0 + n^{-1/2} J^{-1} u)}{\pi(\theta_0)} - 1 \right| \left| \frac{\int_{\Lambda} Z_n(u) du}{\int \tilde{Z}_n(u) du} \right| + \\
&+ \frac{2\bar{c}^{\alpha/2} M_{d,\alpha}^{\alpha/2}}{\int \tilde{Z}_n(u) du} \int_{\Lambda} |Z_n(u) - \tilde{Z}_n(u)| du.
\end{aligned}$$

To bound the very last term we apply Lemma 3 with $c_{d,\alpha} = \bar{c} M_{d,\alpha}/d$ to obtain

$$\frac{2\bar{c}^{\alpha/2} M_{d,\alpha}^{\alpha/2}}{\int \tilde{Z}_n(u) du} \int_{\Lambda} |Z_n(u) - \tilde{Z}_n(u)| du \leq 2\bar{c}^{1+\alpha/2} M_{d,\alpha}^{1+\alpha/2} \lambda_n(c_{d,\alpha}) e^{\bar{c} M_{d,\alpha} \lambda_n(c_{d,\alpha})}$$

which converges to zero under our assumption (iv') .

On the other hand, the first term is bounded by assumption (iv') . Moreover, assumption (iv') also ensures that the term converges to zero as follows

$$\begin{aligned}
2\bar{c}^{\alpha/2} M_{d,\alpha}^{\alpha/2} \sup_{u \in \Lambda} \left| \frac{\pi(\theta_0 + n^{-1/2} J^{-1} u)}{\pi(\theta_0)} - 1 \right| \left| \frac{\int_{\Lambda} Z_n(u) du}{\int \tilde{Z}_n(u) du} \right| &\leq \\
2\bar{c}^{\alpha/2} M_{d,\alpha}^{\alpha/2} e^{\bar{c} M_{d,\alpha} \lambda_n(c_{d,\alpha})} \left| e^{K_n(c_{d,\alpha}) \sqrt{\frac{\|F^{-1}\|_F M_{d,\alpha}}{n}}} - 1 \right| &\rightarrow 0.
\end{aligned}$$

The second term of (C.19) is bounded above by

$$\frac{1}{\int \tilde{Z}_n(u) du} \int_{\Lambda^c} \|u\|^{\alpha} \tilde{Z}_n(u) du + \frac{1}{\int \pi(\theta_0) \tilde{Z}_n(u) du} \int_{\Lambda^c} \|u\|^{\alpha} \pi(\theta_0 + n^{-1/2} J^{-1} u) Z_n(u) du.$$

The first term above converges to zero by standard bounds on gaussian densities for an appropriate choice of the constant \bar{c} (note that \bar{c} can be chosen independently of d and α).

Finally, we bound the last term. Let $\Lambda_k^{\varepsilon} := \{u : \|u\| \in [k\sqrt{\bar{c} M_{d,\alpha}}, (k+1)\sqrt{\bar{c} M_{d,\alpha}}]\}$. Thus we have

$$\int_{\Lambda^c} \|u\|^{\alpha} \pi(\theta_0 + n^{-1/2} J^{-1} u) Z_n(u) du \leq \sum_{k=1}^{\infty} (k+1)^{\alpha} \bar{c}^{\alpha/2} M_{d,\alpha}^{\alpha/2} \int_{\Lambda_k^{\varepsilon}} \pi(\theta_0 + n^{-1/2} J^{-1} u) Z_n(u) du$$

Using Lemma 4 for each integral we have

$$\int_{\Lambda_{\bar{c}}^c} \pi(\theta_0 + n^{-1/2} J^{-1} u) Z_n(u) du \leq \left(\sup_u \pi(u) \right) \left(e^{\bar{c} M_{d,\alpha} \lambda_n(c_{d,\alpha})} \int \tilde{Z}_n(u) du \right) e^{-k \bar{c} M_{d,\alpha} / 8}.$$

Since $M_{d,\alpha} > \max\{1, \alpha\}$ we have

$$\sum_{i=1}^{\infty} (k+1)^\alpha e^{-k \bar{c} M_{d,\alpha} / 8} \leq e^{-\bar{c} M_{d,\alpha} / 10}$$

by choosing \bar{c} large enough. Moreover, our definition of $M_{d,\alpha}$ also implies that

$$\bar{c}^{\alpha/2} M_{d,\alpha}^{\alpha/2} e^{-\bar{c} M_{d,\alpha} / 10} = \exp \left(\frac{\alpha}{2} (\ln \bar{c} + \ln M_{d,\alpha}) - \bar{c} M_{d,\alpha} / 10 \right) \leq \exp(-\bar{c} M_{d,\alpha} / 20)$$

provided that \bar{c} is large enough.

We have that (C.20) can be bounded above by

$$\left(\sup_u \pi(u) \right) \left(e^{\bar{c} M_{d,\alpha} \lambda_n(c_{d,\alpha})} \int \tilde{Z}_n(u) du \right) e^{-\bar{c} M_{d,\alpha} / 20} = o \left(\pi(\theta_0) \int \tilde{Z}_n(u) du \right)$$

under our assumptions. Therefore the result follows. \square

APPENDIX D: PROOFS OF SECTION 3

Proof of Lemma 1. Divide Γ into three regions:

$$(I) := B(0, \bar{k} N_G), \quad (II) := \{\gamma : \max\{\|\gamma\|, \|u_\gamma\|\} \leq \bar{k} N_T\} \setminus B(0, \bar{k} N_G),$$

$$(III) := \Gamma \setminus ((I) \cup (II)),$$

where \bar{k} is chosen later to be large enough independent of the dimensions d or d_1 . Region (I) is defined to be the region where the linear approximation G for $\theta(\cdot)$ is valid in the sense of Assumption A. Region (III) represents the tail of the distribution; either γ or u_γ has large norm. Finally, region (II) is an intermediary region for which G is not a valid approximation but we still have interesting guarantees for deviations from normality. We point out that regions (II) or (III) might be highly non-convex. We will derive sufficient conditions on the values of N_G and N_T as a function of the d and d_1 . It will be sufficient to set $N_G = \sqrt{d}$ and $N_T = \sqrt{d} \log d$.

For notational convenience we define $c_G = \bar{k}^2 N_G^2 / d$ and $c_T = \bar{k}^2 N_T^2 / d$. Our assumptions are such that

$$d \lambda_n(c_G) \rightarrow 0 \quad \text{and} \quad \lambda_n(c_T) < 1/16.$$

We first bound the contribution of region (III) . For any $\gamma \in (III)$, define $\tilde{u}_\gamma = \bar{k} N_G \frac{u_\gamma}{\|u_\gamma\|} \in \mathcal{U}$. Using Lemma 2 we have

$$\ln Z_n(\tilde{u}_\gamma) \leq \langle \Delta_n, \tilde{u}_\gamma \rangle - \frac{1}{2}(1 - 2\lambda_n(c_G))\|\tilde{u}_\gamma\|^2$$

Since $\ln Z_n(\cdot)$ is concave in \mathcal{U} and $\ln Z_n(0) = 0$ by design, we have (D.21)

$$\ln Z_n(u_\gamma) \leq \frac{\|u_\gamma\|}{\|\tilde{u}_\gamma\|} \ln Z_n(\tilde{u}_\gamma) \leq -\|u_\gamma\| N_G \left(\frac{1 - 2\lambda_n(c_G)}{4} \bar{k}^2 \right) \leq -\|u_\gamma\| N_G \frac{\bar{k}^2}{5}$$

by choosing \bar{k} large enough such that $\|\Delta_n\| < \frac{1}{5}\bar{k}N_G$, and using that $\lambda_n(c_G) \leq 1/16$. The contribution of (III) can be bounded by

$$\int_{(III)} \pi(\theta(\eta_0) + n^{-1/2}u_\gamma) Z_n(u_\gamma) d\gamma \leq \pi(\theta_0) \left(\sup_{\theta \in \Theta} \frac{\pi(\theta)}{\pi(\theta_0)} \right) \int_{(III)} \exp\left(-\frac{\bar{k}^2}{5} N_G \|u_\gamma\|\right) d\gamma.$$

The integral on the right can be bounded as follows

$$\begin{aligned} \int_{(III)} \exp\left(-\frac{\bar{k}^2}{5} N_G \|u_\gamma\|\right) d\gamma &\leq \int_{B(0, \bar{k}N_T) \cap (III)} \exp\left(-\frac{\bar{k}^2}{5} N_G \|u_\gamma\|\right) d\gamma \\ &\quad + \int_{B(0, \bar{k}N_T)^c} \exp\left(-\frac{\bar{k}^2}{5} N_G \|u_\gamma\|\right) d\gamma \end{aligned}$$

By definition of (III) , $\gamma \in B(0, N_T) \cap (III)$ implies $\|u_\gamma\| \geq \bar{k}N_T$. On the other hand $\gamma \in B(0, N_T)$ implies that $\|u_\gamma\| \geq \varepsilon_0 N_T$. A standard bound on the integrals yields

$$\begin{aligned} \int_{(III)} \exp\left(-\frac{\bar{k}^2}{5} N_G \|u_\gamma\|\right) d\gamma &\leq \exp\left(-\frac{\bar{k}^3}{5} N_G N_T + d_1 \ln(\bar{k}N_T)\right) + \\ &\quad + \exp\left(-\frac{\bar{k}^2}{5} \varepsilon_0 N_G N_T + d_1 \ln d_1\right). \end{aligned}$$

Using the assumption on the prior, we can bound the contribution of (III) by

$$(D.22) \quad \pi(\theta_0) \exp\left(c_{prior}d + d_1 \ln d_1 + d_1 \ln(\bar{k}N_T) - \frac{\bar{k}^2}{5} \varepsilon_0 N_G N_T\right).$$

Next consider $\gamma \in (II)$. By definition $\gamma \in B(0, \bar{k}N_T) \setminus B(0, \bar{k}N_G)$. Under the assumption that $\lambda_n(c_T) < 1/16$, we have that

$$\ln Z_n(u_\gamma) \leq \langle \Delta_n, u_\gamma \rangle - \frac{1}{2} \frac{7}{8} \|u_\gamma\|^2.$$

Therefore, by choosing \bar{k} such that $\bar{k}N_G > 8\|\Delta_n\|$, we have

$$\begin{aligned} \int_{(II)} \pi(\theta(\eta_0) + n^{-1/2}u_\gamma) Z_n(u_\gamma) d\gamma &\leq \pi(\theta_0) \left(\sup_{\theta \in \Theta} \frac{\pi(\theta)}{\pi(\theta_0)} \right) \int_{(II)} \exp\left(\langle \Delta_n, u_\gamma \rangle - \frac{1}{2} \frac{7}{8} \|u_\gamma\|^2\right) d\gamma \\ &\leq \pi(\theta_0) \left(\sup_{\theta \in \Theta} \frac{\pi(\theta)}{\pi(\theta_0)} \right) \int_{(II)} \exp\left(-\frac{1}{2} \frac{7}{16} \|u_\gamma\|^2\right) d\gamma \end{aligned}$$

Again using our assumption on the prior and standard bounds to gaussian densities, we can bound the contribution of (II) by

$$(D.23) \quad \pi(\theta_0) \exp\left(c_{\text{prior}}d + d_1 \ln(1/\varepsilon_0) - \frac{\varepsilon_0^2}{5} \bar{k}^2 N_G^2\right)$$

Finally, we show a lower bound on the integral over (I) . First note that for any $\gamma \in (I)$ condition (3.10) holds and we have $u_\gamma = r_{1n} + (I + R_{2n})G\gamma$. Therefore, $u_\gamma \in B(0, (\|G\| + \delta_{n2})\bar{k}N_G + \delta_{n1}) \subset B(0, 2\|G\|\bar{k}N_G)$. For simplicity, let $c_{(I)} = 4\|G\|^2 N_G^2/d$.

$$\int_{(I)} \pi\left(\theta(\eta_0) + n^{-1/2}u_\gamma\right) Z_n(u_\gamma) d\gamma \geq \pi(\theta_0) \exp\left(-K_n(c_{(I)})\sqrt{\frac{c_{(I)}d}{n}}\right) \int_{(I)} Z_n(u_\gamma) d\gamma$$

Under our assumptions $\exp\left(-K_n(c_{(I)})\sqrt{\frac{c_{(I)}d}{n}}\right) \rightarrow 1$. Furthermore, using

(3.11), $\|\Delta_n\| = O(\sqrt{d})$, and $\|\gamma\| \leq \bar{k}N_G$, we have

$$\begin{aligned} \ln Z_n(r_{1n} + (I + R_{2n})G\gamma) &= \langle \Delta_n, r_{1n} + (I + R_{2n})G\gamma \rangle - \\ &\quad - \frac{1+2\lambda_n(c_{(I)})}{2} \|r_{1n} + (I + R_{2n})G\gamma\|^2 \\ &\geq o(1) + \langle \Delta_n, G\gamma \rangle - \frac{1+2\lambda_n(c_{(I)})}{2} \|G\gamma\|^2. \end{aligned}$$

Therefore we have

$$\begin{aligned} \int_{(I)} \pi\left(\theta(\eta_0) + n^{-1/2}u_\gamma\right) Z_n(u_\gamma) d\gamma &\geq \pi(\theta_0) O\left(\int_{(I)} \exp\left(\langle \Delta_n, G\gamma \rangle - \frac{1+2\lambda_n(\bar{k}^2)}{2} \|G\gamma\|^2\right) d\gamma\right) \\ &\geq \pi(\theta_0) O\left((1 - 2\lambda_n(c_{(I)}))^{d_1/2} \det(G'G)^{-1/2}\right) \\ &\geq \pi(\theta_0) O(\exp(-\|G\|d_1)). \end{aligned}$$

Choosing $N_G = \sqrt{d}$, $N_T = \sqrt{d} \log d$ and \bar{k} sufficiently large the result follows since we have $d \geq d_1$. ■

Proof of Theorem 3. Let $\hat{\gamma}$ be such that $\hat{\eta} = \eta_0 + n^{1/2}\hat{\gamma}$. We will show that $\ln Z(u_\gamma) < -cd$ for any $\gamma \notin B(0, \bar{k}\sqrt{d})$ where \bar{k} is sufficiently large. Therefore, since the contribution of the prior is bounded by (iv), the MLE $\hat{\gamma} \in B(0, \bar{k}\sqrt{d})$ and the result follows.

Using (D.21) with $N_G = \sqrt{d}$ we have

$$\ln Z(u_\gamma) < -\|u_\gamma\| \sqrt{d} \bar{k}^2/5 \leq -\varepsilon_0 \bar{k}^3 d/5.$$

As stated earlier, the result follows by choosing \bar{k} sufficiently large. ■

Proof of Theorem 4. Using Lemma 1 and known results for gaussian densities, we can restrict our analysis to $B(0, \bar{k}\sqrt{d})$ since the remaining part has negligible mass.

The remaining of the proof follows the same steps in the proof of Theorem 2.

■

REFERENCES

- [1] O. Barndorff-Nielsen (1978). Information and exponential families in statistical theory. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester.
- [2] P. J. Bickel and J. A. Yahav (1969). Some contributions to the asymptotic theory of Bayes solutions, *Z. Wahrsch. Verw. Geb* 11, 257-276.
- [3] G. Chamberlain (1987). Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics* 34, no. 3, 305-334
- [4] V. Chernozhukov and H. Hong (2003). An MCMC approach to classical estimation, *Journal of Econometrics* 115, 293-346.
- [5] S. G. Donald, G. W. Imbens, W. K. Newey (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, 117, no. 1, 55-93.
- [6] R. Dudley (2000). Uniform Central Limit Theorems, Cambridge Studies in advanced mathematics.
- [7] B. Efron (1975). Defining the curvature of a statistical problem (with applications to second order efficiency)(with discussion), *Annals of Statistics*, Vol. 3, No. , 1189-1242.
- [8] B. Efron (1978). The Geometry of Exponential Families, *Annals of Statistics*, Vol. 6, No. 2, 362-376.
- [9] L. P. Hansen and K. J. Singleton (1982). Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, no. 5, 1269-1286.
- [10] I. Ibragimov and R. Has'minskii (1981). Statistical Estimation: Asymptotic Theory, Springer, Berlin.
- [11] G. W. Imbens (1997). One-step estimators for over-identified generalized method of moments models. *Rev. Econom. Stud.* 64, no. 3, 359-383.
- [12] S. Ghosal (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity, *Journal of Multivariate Analysis*, vol 73, 49-68.
- [13] D. Hunter and M. S. Handcock. Inference in curved exponential family models for networks, to appear in *Journal of Graphical and Computational Statistics*,
- [14] E. L. Lehmann and G. Casella (1998). Theory of point estimation. Second edition. Springer Texts in Statistics. Springer-Verlag, New York.
- [15] L. Lovász and S. Vempala, The Geometry of Logconcave Functions and Sampling Algorithms, To appear in *Random Structures and Algorithms*.
- [16] W. K. Newey and D. McFadden (1994). Large Sample Estimation and Hypothesis Testing, *Handbook of Econometrics*, Volume IV, Chapter 36, Edited by R.F. Engle and D.L. McFadden, Elsevier Science.
- [17] S. Portnoy (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* 16, no. 1, 356-366.
- [18] J. Qin and J. Lawless (1994). Empirical Likelihood and General Estimating Equations, *The Annals of Statistics*, Vol. 22, No. 1, 300-325.

E-MAIL: belloni@mit.edu
 1101 KITCHAWAN ROAD,
 BUILDING 800 OFFICE 32-221,
 YORKTOWN HEIGHTS, NY 10598
[HTTP://WEB.MIT.EDU/BELLONI/WWW](http://web.mit.edu/belloni/www)

E-MAIL: vchern@mit.edu
 50 MEMORIAL DRIVE
 ROOM E52-262F
 CAMBRIDGE, MA 02142
[HTTP://WEB.MIT.EDU/VCHERN/WWW](http://web.mit.edu/vchern/www)

Date Due

Lib-26-67

