

---

## Quasi-Maximum Likelihood Estimations (QMLE)

---

The main source for results on QMLE is White (1994). Unfortunately, the book is so technical as to be almost unreadable. We'll try to translate the main results as best we can.

Suppose that  $\{x_t\}, t = 1, \dots, \infty$  is a stochastic process and suppose that we have observed a finite piece of this  $\{x_1, \dots, x_T\}$  and that the true (unknown) log joint density of this can be written

$$\sum_{t=1}^T \log g_t(x_t, \dots, x_1)$$

This is generally no problem for either cross section data (where independence may be a reasonable assumption) or time series models where the data can be thought of as being generated sequentially. Some panel data likelihoods will not, however, be representable in this form.

A (log) quasi likelihood for the data is a collection of density functions indexed by a set of parameters  $\theta$  of the form

$$\sum_{t=1}^T \log f_t(x_t, \dots, x_1; \theta)$$

which it is hoped will include a reasonable approximation to the true density. In practice, this will be the log likelihood for a mathematically convenient representation of the data such as joint Normal. The QMLE is the (or more technically,  $a$ , since there might be non-uniqueness)  $\hat{\theta}$  which maximizes the log quasi-likelihood.

Under the standard types of assumptions which would be used for actual maximum likelihood estimation,  $\hat{\theta}$  proves to be consistent and asymptotically Normal, where the asymptotic distribution is given by  $\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})$ , where  $\mathbf{A}$  is approximated by

$$\mathbf{A}_T = \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 \log f_t}{\partial \theta \partial \theta'}$$

and  $\mathbf{B}$  by (if there is no serial correlation in the gradients)

$$\mathbf{B}_T = \frac{1}{T} \sum_{t=1}^T \left( \frac{\partial \log f_t}{\partial \theta} \right)' \left( \frac{\partial \log f_t}{\partial \theta} \right) \tag{F.1}$$

with the derivatives evaluated at  $\hat{\theta}$ .<sup>1</sup> Serial correlation in the gradients is handled by a Newey-West type calculation in (F.1). This is the standard “sandwich” estimator for the covariance matrix. For instance, if  $\log f_t = -(x_t - z_t\theta)^2$ , (with  $z_t$  treated as exogenous), then

$$\frac{\partial \log f_t}{\partial \theta} = 2(x_t - \theta z_t) z_t'$$

and

$$\frac{\partial^2 \log f_t}{\partial \theta \partial \theta'} = -2z_t' z_t$$

and the asymptotic covariance matrix of  $\hat{\theta}$  is

$$\left( \sum z_t' z_t \right)^{-1} \left( \sum z_t' u_t^2 z_t \right) \left( \sum z_t' z_t \right)^{-1}$$

the standard Eicker-White robust covariance matrix for least squares. Notice that, when you compute the covariance matrix this way, you can be somewhat sloppy with the constant multipliers in the log quasi likelihood—if this were the actual likelihood for a Normal,  $\log f_t$  would have a  $\frac{1}{2\sigma^2}$  multiplier, but that would just cancel out of the calculation since it gets squared in the center factor and inverted in the two ends.

This is very nice, but what is the  $\theta_0$  to which this is converging? After all, nothing above actually required that the  $f_t$  even approximate  $g_t$  well, much less include it as a member. It turns out that this is the value which minimizes the Kullback-Liebler Information Criterion (KLIC) discrepancy between  $f$  and  $g$  which is (suppressing various subscripts) the expected value (over the density  $g$ ) of  $\log(g/f)$ . The KLIC has the properties that it’s non-negative and is equal to zero only if  $f = g$  (almost everywhere), so the QMLE will at least asymptotically come up with the member of the family which is closest (in the KLIC sense) to the truth.

Again, closest might not be close. However, in practice, we’re typically less interested in the complete density function of the data than in some aspects of it, particularly moments. A general result is that if  $f$  is an appropriate selection from the linear exponential family, then the QMLE will provide asymptotically valid estimates of the parameters in a conditional expectation. The linear exponential family are those for which the density takes the form

$$\log f(x; \theta) = a(\theta) + b(x) + \theta' t(x) \tag{F.2}$$

This is a very convenient family because the interaction between the parameters and the data are severely limited.<sup>2</sup> This family includes the Normal,

<sup>1</sup>The formal statement of this requires pre-multiplying the left side by a matrix square root of  $AB^{-1}A$  and having the target covariance matrix be the identity.

<sup>2</sup>The exponential family in general has  $d(\theta)$  entering into that final term, though if  $d$  is invertible, it’s possible to reparameterize to convert a general exponential to the linear form.

gamma (chi-squared and exponential are special cases), Weibull and beta distributions among continuous distributions and binomial, Poisson and geometric among discrete ones. It *does not* include the logistic, t, F, Cauchy and uniform.

For example, suppose that we have “count” data—that is, the observable data are nonnegative integers (number of patents, number of children, number of job offers, etc.). Suppose that we posit that the expected value takes the form  $E(y_t|w_t) = \exp(w_t\theta)$ . The Poisson is a density in the exponential family which has the correct support for the underlying process (that is, it has a positive density only for the non-negative integers). Its probability distribution (as a function of its single parameter  $\lambda$ ) is defined by  $P(x; \lambda) = \frac{\exp(-\lambda)\lambda^x}{x!}$ . If we define  $\omega = \log(\lambda)$ , this is linear exponential family with  $a(\omega) = \exp(\omega)$ ,  $b(x) = \log x!$ ,  $t(x) = x$ . There’s a very good chance that the Poisson will not be the correct distribution for the data because the Poisson has the property that both its mean and its variance are  $\lambda$ . Despite that, the Poisson QMLE, which maximizes  $\sum -\exp(w_t\theta) + x_t(w_t\theta)$ , will give consistent, asymptotically Normal estimates of  $\theta$ .

It can also be shown that, under reasonably general conditions, if the “model” provides a set of moment conditions (depending upon some parameters) that match up with QMLE first order conditions from a linear exponential family, then the QMLE provides consistent estimates of the parameters in the moment conditions.