Description of the ‘ifit‘ algorithm

Guido Masarotto
Università of Padova

Abstract

This brief document details the algorithm implemented in the **ifit** package and provides an overview of the optional arguments for the `ifit::ifit` function.

Keywords: Indirect inference based on intermediate statistics; Generalized method of moments; Parametric estimation; Simulation-based model fitting.

Package **ifit** addresses the problem of fitting a parametric model $\{p_\vartheta \colon \vartheta \in R^p\}$, to some data y_{obs}. It operates under the following assumptions:

1. It is not possible to analytically compute the likelihood function, moments, or other quantities typically used for statistical inference. However, for any given ϑ, it is possible to simulate pseudo-data $y \sim p_\vartheta$.

2. The only prior information available on the parameters are *boundary constraints* where each parameter ϑ_i is known to lie within a specific interval $[l_i, u_i]$, i.e., it is possible to assume that ϑ belongs to the hypercube $\Theta = \{(\vartheta_1, \ldots, \vartheta_p) \in R^p : l_i \leq \vartheta_i \leq u_i\}$.

The implemented estimator is a special case of the class of methods that Jiang and Turnbull (2004) term *indirect inference based on intermediate statistics*. It can also be viewed as an application of the *simulated generalized method of moments* (McFadden 1989; Pakes and Pollard 1989). Additionally, it is closely related to the work of Cox and Kartsonaki (2012) on fitting complex parametric models.

We start defining the score

$$u(y_{obs}; \vartheta) = t_{obs} - \tau(\vartheta),$$

where $t_{obs} = \Psi(y_{obs})$ is a vector of q, $q \geq p$, features intended to summarize the evidence from the data, and $\tau(\vartheta) = E_\vartheta(\Psi(y))$ is its expected value. The package then aims to find

$$\hat{\vartheta} = \arg\min_{\vartheta \in \Theta} u'(y_{obs}; \vartheta) V^{-1} u(y_{obs}; \vartheta)$$

where V is positive-definite weighting matrix.

Assuming the model is well specified (i.e., there exists ϑ_0 such that $y_{obs} \sim p_{\vartheta_0}$), it can be shown that, under the usual regularity and identifiability conditions, the estimates is consistent for every positive-definite matrix V. However, the efficiency depends on V and the most efficient choice for the weighting matrix is $V = \Sigma(\vartheta_0)$, where $\Sigma(\vartheta) = \text{var}_\vartheta(\Psi(y))$. Since ϑ_0 is unknown, this choice of weighting matrix is not directly possible. However, it is easy to proof under the usual assumptions that an asymptotically equivalent estimator to the optimal one can be obtained by solving the quasi-likelihood equation

$$g(\hat{\vartheta}) = J'(\hat{\vartheta})\Sigma^{-1}(\hat{\vartheta})u(y_{obs};\hat{\vartheta}) = 0_p \tag{1}$$

where $J(\vartheta)$ denotes the Jacobian matrix $J(\vartheta) = \partial\tau(\vartheta)/\partial\vartheta$. In addition, continuing to assume that the model is well-specified and that all necessary regularity conditions hold, the dispersion matrix of $\hat{\vartheta}$ can be estimated by $\widehat{\text{var}}(\hat{\vartheta}) = \Omega(\hat{\vartheta})^{-1}$ where $\Omega(\vartheta) = J'(\vartheta)\Sigma^{-1}(\vartheta)J(\vartheta)$.

The implemented algorithm first performs a global search to find a promising starting point. This is followed by a local search, which refines the solution using a trust-region version of the Fisher scoring iteration for solving (1). This basic iteration is defined as

$$\vartheta_{new} = \vartheta_{old} + \Omega(\vartheta)^{-1} g(\vartheta_{old}).$$

Because $\tau(\vartheta)$, $J(\vartheta)$ and $\Omega(\vartheta)$ are unknown, they are approximated through simulations. The approach is sequential. In particular, after step k of the algorithm, N_k pairs $[\vartheta_i; t_i]$, where $t_i = \Psi(y_i)$ with $y_i \sim p_{\vartheta_i}$, are available. These simulated values are then used to decide the location of the next set of simulations, i.e., to choose $\vartheta_{N_k+1}, \ldots, \vartheta_{N_{k+1}}$. When needed, the approximations of $\tau(\vartheta)$, $J(\vartheta)$, and $\Omega(\vartheta)$ are computed from these simulated pairs using local regression techniques. Specifically, a simple kNN average is used to compute the necessary quantities during the global search. In contrast, during the local search, they are obtained by fitting linear models neighborhoods of the current guess including a progressively larger number of points.

The details are as follows.

1: **Input** (default values in brackets): $\texttt{N}_{init}$ (1000), $\texttt{N}_{elite}$ (100), $\texttt{A}_{elite}$ (0.5), $\texttt{Tol}_{global}$ (0.1), $\texttt{NAdd}_{global}$ (100), $\texttt{NTot}_{global}$ (20000), $\texttt{Rho}_{max}$ (0.1), $\texttt{Lambda}$ (0.1), $\texttt{Tol}_{local}$ (1), $\texttt{NFit}_{local}$ (4000), $\texttt{NAdd}_{local}$ (10), $\texttt{Tol}_{model}$ (1.5).

2: Set $k \leftarrow 0$.

3: **Start Global Search**

4: *Initialization.* Set $N_0 \leftarrow \texttt{N}_{init}$ and draw $\vartheta_1, \ldots, \vartheta_{N_0}$ in Θ using Latin hypercube sampling. Simulate the corresponding summary statistics $t_1, \ldots, t_{N_0}$.

5: *Estimation of $\tau(\vartheta)$ and V.* Estimate $\tau(\vartheta)$ using kNN regression. Specifically, for $i = 1, \ldots, N_k$, calculate

$$\hat{\tau}_{k,i} = \sum_{\{r: d_G(\vartheta_r,\vartheta_i) \le \bar{d}_{k,i}\}} \left[1 - \left(\frac{d_G(\vartheta_r, \vartheta_i)}{\bar{d}_{k,i}}\right)^3\right]^3 t_r$$

where $d_G(\vartheta', \vartheta'') = \sqrt{\sum_{i=1}^p [(\vartheta'_i - \vartheta''_i)/(u_i - l_i)]^2}$ and $\bar{d}_{k,i}$ is determined such that the size of the neighbourhood $\{r : d(\vartheta_r, \vartheta_i) \le \bar{d}_{k,i}\}$ is equal to $floor\left(\sqrt{N_k}\right)$. In addition, set $V_k \leftarrow S_k R_k S_k$ where S_k is the $q \times q$ diagonal matrix containing the Median Absolute Deviations (MADs) of the elements of the vector $t_k - \hat{\tau}_{k,i}$ (specifically, one MAD for each summary statistic), and R_k is the correlation matrix among the Gaussian scores (or normal scores) of the same elements of $t_i - \hat{\tau}_{k,i}$ (Boudt, Cornelissen, and Croux 2012).

6: *Elite sample.* Determine the indexes $i_{k,1}, \ldots, i_{k,E_k}$ of the

$$E_k = floor\left[\texttt{N}_{elite} + (\texttt{N}_{init} - \texttt{N}_{elite})\texttt{A}_{elite}^{(N_K/\texttt{N}_{init})^2}\right]$$

couples (ϑ_i, t_i) with the smallest Mahalanobis distances $\big(t_{obs} - \hat{\tau}_{k,i}\big)' V_k^{-1} \big(t_{obs} - \hat{\tau}_{k,i}\big)$. Then, compute

$$M_k \leftarrow \text{ sample mean of } \vartheta_{i_{k,1}}, \ldots, \vartheta_{i_{k,E_k}}$$
$$C_k \leftarrow \text{ sample covariance matrix of } \vartheta_{i_{k,1}}, \ldots, \vartheta_{i_{k,E_k}}$$

and denote with $s_{k,1}, \ldots, s_{k,p}$ the corresponding sample standard deviation, i.e., the square root of the diagonal of C_k.

7: *Convergenge?* If either

(i) $s_{k,r} < \max(1, |M_{k,r}|)\texttt{Tol}_{global}\ \forall r = 1, \ldots, p$, or

(ii) $N_k = \texttt{NTot}_{global}$,

exit from the global search and go to 11.

8: *Elite sample reproduction.* Set $N_{k+1} \leftarrow \min\big(N_k + \texttt{NAdd}_{global}, \texttt{NTot}_{global}\big)$ and sample $\vartheta_{N_K+1}, \ldots, \vartheta_{N_{k+1}}$ from a mixture of E_k multivariate normal distributions (truncated to Θ) with means $\vartheta_{i_{k,1}}, \ldots, \vartheta_{i_{k,E_k}}$ and the same dispersion matrix C_k. Simulate the corresponding summary statistics $t_{N_k+1}, \ldots, t_{N_{k+1}}$.

9: Set $k \leftarrow k + 1$ and go to 5.

10: **End Global Search**

11: Set $L_k \leftarrow \texttt{N}_{elite}$, $\rho_k \leftarrow \texttt{Rho}_{max}/10$ and $\hat{\vartheta}_k \leftarrow \vartheta_{best}$ where ϑ_{best} is the "best point" sampled during the previous phase, i.e., it satisfies

$$\big(t_{obs} - \hat{\tau}_{k,best}\big)' V_k^{-1} \big(t_{obs} - \hat{\tau}_{k,best}\big) = \min_{i=1,\ldots,N_k} \big(t_{obs} - \hat{\tau}_{k,i}\big)' V_k^{-1} \big(t_{obs} - \hat{\tau}_{k,i}\big).$$

12: **Start Local Search**

13: *Estimation of $g(\hat{\vartheta}_k)$, $J(\hat{\vartheta}_k)$ and $\Omega(\hat{\vartheta}_k)$.* Fit to the L_k couples (ϑ_i, t_i) closest to $\hat{\vartheta}_k$, as measured by the distance $d_L(\vartheta, \hat{\vartheta}^{(k)}) = \sum_{i=1}^p \big[(\vartheta_i - \hat{\vartheta}_i^{(k)}) / \max(1, |\vartheta_i^{(k)}|)\big]^2$, the multivariate linear regression model $t_i = \tau + B(\vartheta_i - \hat{\vartheta}_k) + \mathrm{Err}_i$. Denote with $\hat{\tau}_k$, B_k, W_k, H_k the least squares estimates of τ, B, $\mathrm{var}(\mathrm{Err}_i)$ and $\mathrm{var}(\hat{\tau}_k)$, respectively. Further, set

$$J_k \leftarrow \begin{cases} B_k & \text{if } L_k = \texttt{N}_{elite} \\ (1 - \texttt{Lambda}) J_{k-1} + \texttt{Lambda} B_k & \text{otherwise} \end{cases},$$
$$V_k \leftarrow \begin{cases} W_k & \text{if } L_k = \texttt{N}_{elite} \\ (1 - \texttt{Lambda}) V_{k-1} + \texttt{Lambda} W_k & \text{otherwise} \end{cases},$$
$$\Omega_k \leftarrow J_k' V_k^{-1} J_k,$$
$$\hat{g}_k \leftarrow J_k' V_k^{-1} (t_{obs} - \hat{\tau}_k),$$
$$\widehat{\mathrm{var}}(\hat{g}_k) \leftarrow J_k' V_k^{-1} H_k V_k^{-1} J_k.$$

14: *New guess.* Compute $\tilde{\vartheta}_k \leftarrow \hat{\vartheta}_k + \delta_k$ where δ_k is the solution of the linear programming problem

$$\min_{\delta \in \Delta_k} \sum_{i=1}^p |r_{k,i}(\delta)|$$

where $r_k(\delta) = \Omega_k \delta - \hat{g}_k$ and $\Delta_k = \big\{\delta \in R^p : \hat{\vartheta}_k + \delta \in \Theta \text{ and } |\delta_i| \leq \max(1, |\hat{\vartheta}_{k,i}|)\rho_k\big\}$.

15: *Convergence?* If $L_k = \texttt{NFit}_{local}$ and $\hat{g}_k'\widehat{\text{var}}(\hat{g}_k)^{-1}\hat{g}_k < p\texttt{Tol}_{local}$ exit from the local search and go to 20.

16: *Sampling near the new guess.* Set $N_{k+1} \leftarrow N_k + \texttt{NAdd}_{local}$ and draw $\vartheta_{N_k+1}, \ldots, \vartheta_{N_{k+1}}$ uniformly in $\{\vartheta \in \Theta : (\vartheta - \tilde{\vartheta}_k)'\Omega_k(\vartheta - \tilde{\vartheta}_k) \leq 1\}$. Simulate the corresponding summary statistics $t_{N_k+1}, \ldots, t_{N_{k+1}}$.

17: *Accept/reject the guess. Adjust the size of the trust region.* Compute the differences between the last sampled summary statistics and their predictions obtained from the current linear model $D_i = t_i - \hat{\tau}_k - B_k(\vartheta_i - \tilde{\vartheta}_k)$ for $i = N_k + 1, \ldots, N_{k+1}$. If

$$\sum_{i=N_k+1}^{N_{k+1}} D_i' V_k^{-1} D_i < q(N_{k+1} - N_k)Mod_{ok}$$

accept the proposal and set $\hat{\vartheta}_{k+1} \leftarrow \tilde{\vartheta}_k$ and $\rho_{k+1} \leftarrow \min(2\rho_k, \texttt{Rho}_{max})$. Otherwise, set $\hat{\vartheta}_{k+1} \leftarrow \hat{\vartheta}_k$ and $\rho_{k+1} \leftarrow \rho_k/4$.

18: Set $k \leftarrow k + 1$, $L_k \leftarrow \min(\texttt{NFit}_{local}, L_{k-1} + \texttt{NAdd}_{local})$ and go to 13.

19: **End Local Search**

20: Set $\hat{\vartheta} \leftarrow \tilde{\vartheta}_k$ and $\widehat{\text{var}}(\hat{\vartheta}) = \Omega_k^{-1}$.

References

Boudt K, Cornelissen J, Croux C (2012). “The Gaussian Rank Correlation Estimator: Robustness Properties.” *Statistics and Computing*, **22**(2), 471–483. ISSN 1573-1375. doi:10.1007/s11222-011-9237-0.

Cox DR, Kartsonaki C (2012). “The Fitting of Complex Parametric Models.” *Biometrika*, **99**(3), 741–747. doi:10.1093/biomet/ass030.

Jiang W, Turnbull B (2004). “The Indirect Method: Inference Based on Intermediate Statistics: A Synthesis and Examples.” *Statistical Science*, **19**(2), 239–263. doi:10.1214/088342304000000152.

McFadden D (1989). “A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration.” *Econometrica*, **57**(5), 995–1026. doi:10.2307/1913621.

Pakes A, Pollard D (1989). “Simulation and the Asymptotics of Optimization Estimators.” *Econometrica*, **57**(5), 1027–1057. doi:10.2307/1913622.

Affiliation:

Guido Masarotto
Department of Statistical Sciences
University of Padova, Italy
Email: guido.masarotto@unipd.it