Linear aggregation of frontier estimators

Rao Salim

Linear aggregation of frontier estimators

Salim Rao

Bengal Engineering and Science University, Shibpur

Abstract

In this paper, we propose a linear aggregation technique to estimate the frontier of a dataset. Our approach is based on recent techniques of Conditional Extreme Value Analysis.

Let X be a random variable and $p(x)$ its probability density function, $F(x) = P(X \leq x)$ its distribution function, and $\Psi(x) = 1-F(x)$ its complementary distribution function. We can define a new random variable, Y_n, as the maximum of n copies of the random variable X: $Y_n = \max\{X_1, X_2, ..., X_n\}$. Y_n is the n-sample maximum of the random variable X. If the events generating the realizations of X are independent, the cumulative distribution of Y_n may be expressed as $[F(y)]^n$. Upon definition of a renormalized variable $S_n = (Y_n - b_n)/a_n$, the extreme value theorem establishes that

Theorem *If*

$$\lim_{n\to\infty} P(S_n < s) = \lim_{n\to\infty} F^n(a_n s + b_n) = H(s) \tag{1}$$

where $a_n > 0$ and b_n are normalization constants, then the function $H(s)$ in Eq. (1) must be one of the three following types:

- *EV1 or Gumbel:* $H(s) = \exp(-\exp(-s))$
- *EV2 or Fréchet:* $H(s) = \exp(-s^{-\alpha})$
- *EV3 or Weibull:* $H(s) = \exp(-|s|^{\alpha})$

The three asymptotic types, EV1-EV3, can be thought of as special cases of a single Generalized Extreme Value distribution (GEV) :

$$H_{GEV}(s) = \exp\left\{ -\left(1 + \gamma\frac{s-\mu}{\sigma}\right)_{+}^{-1/\gamma} \right\} \tag{2}$$

where $(.)_{+} = \max(.,0)$, μ is the location parameter, $\sigma > 0$ is the scale parameter, and γ is a shape parameter. The limit $\gamma = 0$ corresponds to the EV1 distribution, $\gamma > 0$ to the EV2 distribution (with $\alpha = 1/\gamma$) and $\gamma < 0$ to the EV3 distribution (with $\alpha = -1/\gamma$). The function $H_{GEV}(s)$ is usually fitted to the cumulative distribution of non-normalized maxima, so that the location parameter μ and the scale parameter σ are the renormalization parameters b_n and a_n respectively. However, it is important to note that the distribution describing the n-sample maximum will strictly be a GEV only for large values of n. How large the value of n needs to be should be determined by analyzing the convergence properties based on the observed realizations of X.

When the studied phenomena depends on a covariate, one has to deal with Conditional Extreme Value Analysis (CEVA). This branch of statistics has become very active these past ten years, the main contributions to this domain are listed below:

- Theoretical issues: [65, 63, 64, 26, 32, 20, 31, 33, 30, 5, 27, 13, 51, 25, 4, 1, 82, 81]
- Quantile regression: [56, 19, 66]
- Application to finance: [10, 3, 35, 11, 58, 80]

Edge estimation [53, 52], frontier estimation [62, 78, 75, 77, 72, 16, 14, 15, 9, 2, 39, 8, 61, 17, 70, 69, 21, 79, 60, 22, 48, 36, 59, 23, 24, 68, 6, 73, 12, 55, 57, 67, 54, 43, 45, 18, 29, 71, 74, 76, 38] and boundary estimation [44, 47, 46, 42, 41, 40, 49, 28, 50, 7, 34, 37] are particular cases of CEVA. They are embedded in the situation where the conditional extreme-value index is negative.

Our mail result is the following:

Theorem *Let $\hat{\phi}_1(x), \ldots, \hat{\phi}_K(x)$ be K consistent estimators of a frontier $\phi(x)$. Let $w_1, \ldots, w_K$ be K weights summing to one. Then,*

$$\hat{\phi}(x) := \sum_{i=1}^{K} w_i \hat{\phi}_i(x)$$

is a consistent estimator of the frontier $\phi(x)$.

Proof. The proof is based on the continuity of the sum function and on the property that $\sum_{i=1}^{K} w_i = 1$. Let us highlight that the positivity of the weights in not required.

References

[1] B. Abdous, A.L. Fougères, and K. Ghoudi. Extreme behaviour for bivariate elliptical distributions. *Canadian Journal of Statistics*, 33(3):317–334, 2005.

[2] Y. Aragon, A. Daouia, and C. Thomas-Agnan. Nonparametric frontier estimation: A conditional quantile-based approach. *Econometric Theory*, 21(2):358–389, 2005.

[3] T.G. Bali and D. Weinbaum. A conditional extreme value volatility estimator based on high-frequency returns. *Journal of Economic Dynamics and Control*, 31(2):361–397, 2007.

[4] J. Beirlant, T. De Wet, and Y. Goegebeur. Nonparametric estimation of extreme conditional quantiles. *Journal of statistical computation and simulation*, 74(8):567–580, 2004.

[5] J. Beirlant and Y. Goegebeur. Local polynomial maximum likelihood estimation for pareto-type distributions. *Journal of Multivariate Analysis*, 89(1):97–118, 2004.

[6] V.E. Berezkin, G.K. Kamenev, and A.V. Lotov. Hybrid adaptive methods for approximating a nonconvex multidimensional pareto frontier. *Computational Mathematics and Mathematical Physics*, 46(11):1918–1931, 2006.

[7] H. Bi, G. Wan, and N.D. Turvey. Estimating the self-thinning boundary line as a density-dependent stochastic biomass frontier. *Ecology*, 81(6):1477–1483, 2000.

[8] G. Bouchard, S. Girard, A. Iouditski, and A. Nazin. Some linear programming methods for frontier estimation. *Applied Stochastic Models in Business and Industry*, 21(2):175–185, 2005.

[9] G. Bouchard, S. Girard, AB Iouditski, and AV Nazin. Nonparametric frontier estimation by linear programming. *Automation and Remote Control*, 65(1):58–64, 2004.

[10] H.N.E. Byström. Managing extreme risks in tranquil and volatile markets using conditional extreme value theory. *International Review of Financial Analysis*, 13(2):133–152, 2004.

[11] H.N.E. Byström. Extreme value theory and extremely large electricity price changes. *International Review of Economics & Finance*, 14(1):41–55, 2005.

[12] C. Cazals, J.P. Florens, and L. Simar. Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106(1):1–25, 2002.

[13] V. Chavez-Demoulin and A.C. Davison. Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):207–222, 2005.

[14] A. Daouia. Asymptotic representation theory for nonstandard conditional quantiles. *Journal of Nonparametric Statistics*, 17(2):253–268, 2005.

[15] A. Daouia, J.P. Florens, and L. Simar. Functional convergence of quantile-type frontiers with application to parametric approximations. *Journal of Statistical Planning and Inference*, 138(3):708–725, 2008.

[16] A. Daouia, J.P. Florens, and L. Simar. Frontier estimation and extreme value theory. *Bernoulli*, 16(4):1039–1063, 2010.

[17] A. Daouia, J.P. Florens, and L. Simar. Regularization of nonparametric frontier estimators. *Journal of Econometrics*, 2012.

[18] A. Daouia, L. Gardes, and S. Girard. Nadarayas estimates for large quantiles and free disposal support curves. *Exploring Research Frontiers in Contemporary Statistics and Econometrics*, pages 1–22, 2012.

[19] A. Daouia, L. Gardes, and S. Girard. On kernel smoothing for extremal quantile regression. *Bernoulli*, 2012. DOI: 10.3150/12-BEJ466.

[20] A. Daouia, L. Gardes, S. Girard, and A. Lekina. Kernel estimators of extreme level curves. *Test*, 20(2):311–333, 2011.

[21] A. Daouia and I. Gijbels. Robustness and inference in nonparametric partial frontier modeling. *Journal of Econometrics*, 161(2):147–165, 2011.

[22] A. Daouia and I. Gijbels. Estimating frontier cost models using extremiles. *Exploring Research Frontiers in Contemporary Statistics and Econometrics*, pages 65–81, 2012.

[23] A. Daouia and A. Ruiz-Gazen. Robust nonparametric frontier estimators: qualitative robustness and influence function. *Statistica Sinica*, 16(4):1233, 2006.

[24] A. Daouia and L. Simar. Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics*, 140(2):375–400, 2007.

[25] B. Das and S.I. Resnick. Conditioning on an extreme component: Model consistency with regular variation on cones. *Bernoulli*, 17(1):226–252, 2011.

[26] B. Das and S.I. Resnick. Detecting a conditional extreme value model. *Extremes*, 14(1):29–61, 2011.

[27] AC Davison and NI Ramesh. Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):191–208, 2000.

[28] A. Delaigle and I. Gijbels. Estimation of boundary and discontinuity points in deconvolution problems. *Statistica Sinica*, 16(3):773, 2006.

[29] J.P. Florens and L. Simar. Parametric approximations of nonparametric frontiers. *Journal of Econometrics*, 124(1):91–116, 2005.

[30] L. Gardes and S. Girard. A moving window approach for nonparametric estimation of the conditional tail index. *Journal of Multivariate Analysis*, 99(10):2368–2388, 2008.

[31] L. Gardes and S. Girard. Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels. *Extremes*, 13(2):177–204, 2010.

[32] L. Gardes and S. Girard. Functional kernel estimators of large conditional quantiles. *Electronic Journal of Statistics*, 6:1715–1744, 2012.

[33] L. Gardes, S. Girard, and A. Lekina. Functional nonparametric estimation of conditional extreme quantiles. *Journal of Multivariate Analysis*, 101(2):419–433, 2010.

[34] J. Geffroy, S. Girard, and P. Jacob. Asymptotic normality of the l 1-error of a boundary estimator. *Journal of Nonparametric Statistics*, 18(1):21–31, 2006.

[35] A. Ghorbel and A. Trabelsi. Predictive performance of conditional extreme value theory in value-at-risk estimation. *International Journal of Monetary Economics and Finance*, 1(2):121–148, 2008.

[36] I. Gijbels, E. Mammen, B.U. Park, and L. Simar. On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association*, 94(445):220–228, 1999.

[37] S. Girard. On the asymptotic normality of the l1-error for haar series estimates of poisson point processes boundaries. *Statistics & Probability Letters*, 66(1):81–90, 2004.

[38] S. Girard, A. Guillou, and G. Stupfler. Frontier estimation with kernel regression on high order moments. *Journal of Multivariate Analysis*, 116:172–189, 2013.

[39] S. Girard, A. Iouditski, and A.V. Nazin. L 1-optimal nonparametric frontier estimation via linear programming. *Automation and Remote Control*, 66(12):2000–2018, 2005.

[40] S. Girard and P. Jacob. Extreme values and haar series estimates of point process boundaries. *Scandinavian Journal of Statistics*, 30(2):369–384, 2003.

[41] S. Girard and P. Jacob. Projection estimates of point processes boundaries. *Journal of Statistical Planning and Inference*, 116(1):1–15, 2003.

[42] S. Girard and P. Jacob. Extreme values and kernel estimates of point processes boundaries. *ESAIM: Probability and Statistics*, 8(1):150–168, 2004.

[43] S. Girard and P. Jacob. Frontier estimation via kernel regression on high power-transformed data. *Journal of Multivariate Analysis*, 99(3):403–420, 2008.

[44] S. Girard and P. Jacob. A note on extreme values and kernel estimators of sample boundaries. *Statistics & Probability Letters*, 78(12):1634–1638, 2008.

[45] S. Girard and P. Jacob. Frontier estimation with local polynomials and high power-transformed data. *Journal of Multivariate Analysis*, 100(8):1691–1705, 2009.

[46] S. Girard and L. Menneteau. Central limit theorems for smoothed extreme value estimates of poisson point processes boundaries. *Journal of Statistical Planning and Inference*, 135(2):433–460, 2005.

[47] S. Girard and L. Menneteau. Smoothed extreme value estimators of non-uniform point processes boundaries with application to star-shaped supports estimation. *Communications in StatisticsTheory and Methods*, 37(6):881–897, 2008.

[48] W.H. Greene. Maximum likelihood estimation of econometric frontier functions. *Journal of Econometrics*, 13(1):27–56, 1980.

[49] P. Hall, L. Peng, and C. Rau. Local likelihood tracking of fault lines and boundaries. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):569–582, 2001.

[50] P. Hall and L. Simar. Estimating a changepoint, boundary, or frontier in the presence of observation error. *Journal of the American Statistical Association*, 97(458):523–534, 2002.

[51] P. Hall and N. Tajvidi. Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science*, pages 153–167, 2000.

[52] P. Jacob and C. Suquet. Estimating the edge of a poisson process by orthogonal series. *Journal of Statistical Planning and Inference*, 46(2):215–234, 1995.

[53] P. Jacob and C. Suquet. Regression and edge estimation. *Statistics & Probability Letters*, 27(1):11–15, 1996.

[54] S.O. Jeong and L. Simar. Linearly interpolated fdh efficiency score for nonconvex frontiers. *Journal of Multivariate Analysis*, 97(10):2141–2161, 2006.

[55] A. Kneip, L. Simar, and P.W. Wilson. Asymptotics for dea estimators in nonparametric frontier models. Technical report, Discussion paper, 2003.

[56] R. Koenker and K. Hallock. Quantile regression: An introduction. *Journal of Economic Perspectives*, 15(4):43–56, 2001.

[57] S.C. Kumbhakar and C.A.K. Lovell. *Stochastic frontier analysis*. Cambridge University Press, 2003.

[58] V. Marimoutou, B. Raggad, and A. Trabelsi. Extreme value theory and value at risk: application to oil market. *Energy Economics*, 31(4):519–530, 2009.

[59] C. Martins-Filho and F. Yao. A smooth nonparametric conditional quantile frontier estimator. *Journal of Econometrics*, 143(2):317–333, 2008.

[60] B.U. Park and L. Simar. Efficient semiparametric estimation in a stochastic frontier model. *Journal of the American Statistical Association*, 89(427):929–936, 1994.

[61] L. Peng. Bias-corrected estimators for monotone and concave frontier functions. *Journal of Statistical Planning and Inference*, 119(2):263–275, 2004.

[62] S. Rao. Frontier estimation as a particular case of conditional extreme value analysis. http://hal.archives-ouvertes.fr/hal-00776107, 2013.

[63] S. Rao. Linear aggregation of conditional extreme-value index estimators. http://hal.archives-ouvertes.fr/hal-00771647, 2013.

[64] S. Rao. Nonlinear aggregation of conditional extreme-value index estimators. http://hal.archives-ouvertes.fr/hal-00776108, 2013.

[65] S. Rao. A review on conditional extreme value analysis. http://hal.archives-ouvertes.fr/hal-00770546, 2013.

[66] O. Rosen and A. Cohen. Extreme percentile regression. In *Statistical Theory and Computational Aspects of Smoothing: Proceedings of the COMPSTAT94 satellite meeting held in Semmering, Austria*, pages 27–28, 1994.

[67] P. Schmidt and T.F. Lin. Simple tests of alternative specifications in stochastic frontier models. *Journal of Econometrics*, 24(3):349–361, 1984.

[68] J.K. Sengupta. Stochastic data envelopment analysis: a new approach. *Applied Economics Letters*, 5(5):287–290, 1998.

[69] L. Simar. Aspects of statistical analysis in dea-type frontier models. *Journal of Productivity Analysis*, 7(2):177–185, 1996.

[70] L. Simar. Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis*, 20(3):391–424, 2003.

[71] L. Simar and P.W. Wilson. Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management science*, 44(1):49–61, 1998.

[72] L. Simar and P.W. Wilson. Of course we can bootstrap dea scores! but does it mean anything? logic trumps wishful thinking. *Journal of Productivity Analysis*, 11(1):93–97, 1999.

[73] L. Simar and P.W. Wilson. A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics*, 27(6):779–802, 2000.

[74] L. Simar and P.W. Wilson. Statistical inference in nonparametric frontier models: The state of the art. *Journal of Productivity Analysis*, 13(1):49–78, 2000.

[75] L. Simar and P.W. Wilson. Testing restrictions in nonparametric efficiency models. *Communications in Statistics-Simulation and Computation*, 30(1):159–184, 2001.

[76] L. Simar and P.W. Wilson. Statistical inference in nonparametric frontier models: recent developments and perspectives. *The measurement of productive efficiency and productivity growth*, pages 421–521, 2008.

[77] L. Simar and P.W. Wilson. Inferences from cross-sectional, stochastic frontier models. *Econometric Reviews*, 29(1):62–98, 2009.

[78] L. Simar and P.W. Wilson. Performance of the bootstrap for dea estimators and iterating the principle. *Handbook on data envelopment analysis*, pages 241–271, 2011.

[79] L. Simar and V. Zelenyuk. Stochastic fdh/dea estimators for frontier analysis. *Journal of Productivity Analysis*, 36(1):1–20, 2011.

[80] C. Wang, H. Zhuang, Z. Fang, and T. Lu. A model of conditional var of high frequency extreme value based on generalized extreme value distribution. *Journal of Systems & Management*, 3:003, 2008.

[81] H. Wang and C.L. Tsai. Tail index regression. *Journal of the American Statistical Association*, 104(487):1233–1240, 2009.

[82] H.J. Wang, D. Li, and X. He. Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, (to appear), 2012.