

## Data-driven wavelet-Fisz methodology for nonparametric function estimation

Piotr Z. Fryzlewicz, *University of Bristol*

### Abstract

We propose a wavelet-based technique for the nonparametric estimation of functions contaminated with noise whose mean and variance are linked via a possibly unknown variance function. Our method, termed the data-driven wavelet-Fisz technique, consists of estimating the variance function via a Nadaraya-Watson estimator, and then performing a wavelet thresholding procedure which uses the estimated variance function and local means of the data to set the thresholds at a suitable level. We demonstrate the mean-square near-optimality of our wavelet estimator over the usual range of Besov classes. To achieve this, we establish an exponential inequality for the Nadaraya-Watson variance function estimator. We discuss various implementation issues concerning our wavelet estimator, and demonstrate its good practical performance. We also show how it leads to a new wavelet-domain data-driven variance-stabilising transform. Our estimator can be applied to a variety of problems, including the estimation of volatilities, spectral densities and Poisson intensities, as well as to a range of problems in which the distribution of the noise is unknown.

AMS 2000 subject classifications: Primary 62G08; secondary 62G05, 62G20.

Keywords: Besov spaces, exponential inequality, heteroscedasticity, Nadaraya-Watson estimator, nonparametric regression, variance function, variance-stabilising transform, wavelets.



Full Text: [PDF](#)

Fryzlewicz, Piotr Z., Data-driven wavelet-Fisz methodology for nonparametric function estimation, *Electronic Journal of Statistics*, 2, (2008), 863-896 (electronic). DOI: 10.1214/07-EJS139.

## References

- F. Abramovich, Y. Benjamini, D. Donoho, and I. Johnstone. Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.*, 34: 584–653, 2006. [MR2281879](#)
- A. Antoniadis, P. Besbeas, and T. Sapatinas. Wavelet shrinkage for natural exponential families with cubic variance functions. *Sankhya Ser. A*, 63: 309–327, 2001. [MR1897045](#)
- A. Antoniadis and T. Sapatinas. Wavelet shrinkage for natural exponential families with quadratic variance functions. *Biometrika*, 88: 805–820, 2001. [MR1859411](#)
- K.J. Archer. Graphical technique for identifying a monotonic variance stabilizing transformation for absolute gene intensity signals. *BMC Bioinformatics*, 5: 60, 2004.
- P. Besbeas, I. De Feis, and T. Sapatinas. A comparative study of wavelet shrinkage estimators for Poisson counts. *Int. Statist. Review*, 72: 209–237, 2004.
- L. Breiman and J.H. Friedman. Estimating optimal transformations for multiple regression and correlation. *J. Am. Statist. Ass.*, 80: 580–619, 1985. [MR0803258](#)

- L.D. Brown, T. Cai, R. Zhang, L. Zhao, and H. Zhou. The root-unroot algorithm for density estimation as implemented via wavelet block thresholding. Preprint, 2007.
- T. Cai and L. Wang. Adaptive variance function estimation in heteroscedastic nonparametric regression. *Ann. Stat.*, to appear, 2007.
- J-M. Chiou and H-G. Müller. Nonparametric quasi-likelihood. *Ann. Stat.*, 27: 36–64, 1999. [MR1701100](#)
- H. Dette and K. Pilz. On the estimation of a monotone conditional variance in nonparametric regression. *Ann. Inst. Stat. Math.*, to appear, 2007.
- D. Donoho and I. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, 90: 1200–1224, 1995. [MR1379464](#)
- D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81: 425–455, 1994. [MR1311089](#)
- M. Fisz. The limiting distribution of a function of two independent random variables and its statistical application. *Colloquium Mathematicum*, 3: 138–146, 1955. [MR0069412](#)
- P. Fryzlewicz, V. Delouille, and G. Nason. GOES-8 X-ray sensor variance stabilization using the multiscale data-driven Haar-Fisz transform. *J. Roy. Statist. Soc. C*, 56: 99–116, 2007. [MR2339165](#)
- P. Fryzlewicz and G. P. Nason. A Haar-Fisz algorithm for Poisson intensity estimation. *Journal of Computational and Graphical Statistics*, 13: 621–638, 2004. [MR2087718](#)
- P. Fryzlewicz, G.P. Nason, and R. von Sachs. A wavelet-Fisz approach to spectrum estimation. *J. Time Ser. Anal.*, 29: 868–880, 2008.
- P. Fryzlewicz, T. Sapatinas, and S. Subba Rao. A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika*, 93: 687–704, 2006. [MR2261451](#)
- H-Y. Gao. Choice of thresholds for wavelet shrinkage estimate of the spectrum. *J. Time Ser. Anal.*, 18: 231–252, 1997a. [MR1456639](#)
- H-Y. Gao. Wavelet shrinkage estimates for heteroscedastic regression models. *Statistical Sciences Division, MathSoft, Inc.*, 1997b.
- T. Gasser, A. Kneip, and W. Koehler. A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.*, 86: 643–652, 1991. [MR1147088](#)
- M. Jansen. Multiscale Poisson data smoothing. *J. R. Statist. Soc. B*, 68: 27–48, 2006. [MR2212573](#)
- M. Jansen, G.P. Nason, and B.W. Silverman. Multiscale methods for data on graphs and irregular multidimensional situations. *J. R. Statist. Soc. Series B*, to appear, 71, 2009.
- W.B. Johnson, G. Schechtman, and J. Zinn. Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Prob.*, 13: 234–253, 1985. [MR0770640](#)
- I. Johnstone and B. Silverman. Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. Ser. B*, 59: 319–351, 1997. [MR1440585](#)
- I. Johnstone and B. Silverman. Empirical Bayes selection of wavelet thresholds. *Ann. Statist.*, 33: 1700–1752, 2005a. [MR2166560](#)
- I.M. Johnstone and B.W. Silverman. EBayesThresh: R and S-Plus programs for Empirical Bayes thresholding. *J. Statist. Software*, 12.8: 1–38, 2005b.
- C. Joutard. Sharp large deviations in nonparametric estimation. *J. Nonpar. Statist.*, 18: 293–306, 2006. [MR2263101](#)

- E. Kolaczyk and R. Nowak. Multiscale likelihood analysis and complexity penalized estimation. *Ann. Statist.*, 32:500–527, 2004. [MR2060167](#)
- E. D. Kolaczyk. Bayesian multiscale models for Poisson processes. *Journal of the American Statistical Association*, 94:920–933, 1999. [MR1723303](#)
- O.B. Linton, R. Chen, N.S. Wang, and W. Härdle. An analysis of transformations for additive nonparametric regression. *J. Am. Statist. Ass.*, 92:1512–1521, 1997. [MR1615261](#)
- D. Louani. Some large deviations limit theorems in conditional nonparametric statistics. *Statistics*, 33:171–196, 1999. [MR1744752](#)
- S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattn Anal. Mach. Intell.*, 11:674–693, 1989.
- Y. Meyer. *Wavelets and Operators*. Cambridge University Press, 1992. [MR1228209](#)
- E. Motakis, G. Nason, P. Fryzlewicz, and G. Rutter. Variance stabilization and normalization for one-color microarray data using a data-driven multiscale approach. *Bioinformatics*, 22:2547–2553, 2006.
- P. Moulin. Wavelet thresholding techniques for power spectrum estimation. *IEEE Trans. Sig. Proc.*, 42:3126–3136, 1994.
- G. P. Nason and B. W. Silverman. The stationary wavelet transform and some statistical applications. In A. Antoniadis and G. Oppenheim, editors, *Lecture Notes in Statistics*, vol. 103, pages 281–300. Springer-Verlag, 1995.
- M. H. Neumann. Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *Journal of Time Series Analysis*, 17: 601–633, 1996. [MR1424908](#)
- M. Pensky, B. Vidakovic, and D. de Canditiis. Bayesian decision theoretic scale-adaptive estimation of a log-spectral density. *Stat. Sinica*, 17:635–666, 2007.
- D.M. Rocke and B.P. Durbin. A model for measurement error for gene expression arrays. *J. Comput. Biol.*, 8:557–569, 2001.
- H.P. Rosenthal. On the subspaces of  $\ell^p$  ( $p > 2$ ) spanned by sequences of independent random variables. *Israel J. Math.*, 8:273–303, 1970. [MR0271721](#)
- R. Rudzakis, L. Saulis, and V. Statulevicius. A general lemma on probabilities of large deviations. *Lithuanian Math. J.*, 18:226–238, 1978. [MR0501287](#)
- S. Sardy, A. Antoniadis, and P. Tseng. Automatic smoothing with wavelets for a wide class of distributions. *Journal of Computational and Graphical Statistics*, 13:399–421, 2004. [MR2063992](#)
- R. Tibshirani. Estimating transformations for regression via additivity and variance stabilization. *J. Am. Stat. Ass.*, 83:394–405, 1988. [MR0971365](#)
- K. E. Timmermann and R. D. Nowak. Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging. *IEEE Transactions on Information Theory*, 45:846–862, 1999. [MR1682515](#)
- B. Vidakovic. *Statistical Modeling by Wavelets*. Wiley, New York, 1999. [MR1681904](#)
- R. von Sachs and B. MacGibbon. Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scand. J. Statist.*, 27: 475–499, 2000. [MR1795776](#)
- L. Wang, L. Brown, T. Cai, and M. Levine. Effect of mean on variance function estimation in nonparametric regression. *Ann. Statist.*, 36:646–664, 2008. [MR2396810](#)

