MR4939548 62M10 62G08 62M20

Wei, James L. (4-ICL-M); **Nason, Guy P.** (4-ICL-M)

Fully nonparametric MIDAS: a new approach for nonparametric mixed frequency time series regression. (English. English summary)

Electron. J. Stat. **19** (2025), *no.* 2, 3292–3316.

Mixed data sampling (MIDAS) regression models allows us to use the information contained in high-frequency regressors to predict low-frequency variables. In this article, a new class of MIDAS regression models, called the fully nonparametric mixed data sampling (FNP-MIDAS) regression model, is proposed. To obtain an h-period ahead forecast of a low-frequency variable y_t^L using N high-frequency predictors $x_{1,t}^H, \dots, x_{N,t}^H$, the authors define the FNP-MIDAS regression model in an additive form as

$$y_{t+h}^L = \alpha_h + \sum_{i=1}^{N} \sum_{j=0}^{J_i} \beta_{h,i,j} \phi_{h,i} \left(x_{i,t-j/m}^H \right) + \epsilon_{t+h}^L,$$

where J_i is the maximum lag order of the i-th predictor $x_{i,t}^H$, $\phi_{h,i}(\cdot)$ is the component function that transforms this predictor regardless of its lag orders, and ϵ_t^L is an *i.i.d.* error term. This model can be viewed as an extension of nonparametric MIDAS by J. Breitung and C. Roling [J. Forecast. **34** (2015), no. 7, 588–603; MR3416308] and an alternative to semiparametric MIDAS by E. Ghysels, V. Kvedaras and V. Zemlys-Balevičius [in *Handbook of statistics. Vol. 42*, 117–153, Elsevier, Amsterdam, 2020, doi:10.1016/bs.host.2019.01.005].

FNP-MIDAS is primarily designed to improve forecasting performance. From this viewpoint, the main focus of this article is on presenting algorithms to estimate N component functions $\{\phi_{h,i}\}_{i=1}^N$ and $\sum_{i=1}^N (J_i+1)+1$ coefficients α_h, $\{\{\beta_{h,i,j}\}_{j=0}^{J_i}\}_{i=1}^N$. These parameters can be estimated by a simple backfitting algorithm that alternates coefficient estimation and component function estimation at each iteration (whereas their joint estimation is not possible because of identification issues). Convergence of the estimation algorithms and uniqueness of component function estimates are demonstrated. It is of second interest to estimate the underlying data generating process by FNP-MIDAS, and thus consistency of the estimators is not proven. *Masayuki Hirukawa*

[References]

1. Abhyankar, A., Copeland, L. S. and Wong, W. (1997). Uncovering nonlinear structure in real-time stock-market indexes: the S&P 500, the DAX, the Nikkei 225, and the FTSE-100. *Journal of Business & Economic Statistics* **15** 1–14.
2. Aiken, E. L., Nguyen, A. T., Viboud, C. and Santillana, M. (2021). Toward the use of neural networks for influenza prediction at multiple spatial resolutions. *Science Advances* **7** eabb1237.
3. Andreou, E., Ghysels, E. and Kourtellos, A. (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics* **158** 246–261. MR2720834 MR2720834
4. Andreou, E., Ghysels, E. and Kourtellos, A. (2011). *Oxford Handbook of Economic Forecasting* Forecasting with mixed-frequency data, 225–245. Oxford University Press. MR3204194 MR3204194
5. Armesto, M. T., Engemann, K. M. and Owyang, M. T. (2010). Forecasting with Mixed Frequencies. *Federal Reserve Bank of St Louis Review* **92** 521–536.
6. Arnold, T. B. and Tibshirani, R. J. (2020). Introduction to the genlasso package. *URL* https://CRAN.R-Project.org/package=genlasso. *Vignette included in R package genlasso, version 1.5.*

7. Baquero, O. S., Santana, L. M. R. and Chiaravalloti-Neto, F. (2018). Dengue forecasting in São Paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models. *PLoS One* **13** e0195065.
8. Barmpadimos, I., Hueglin, C., Keller, J., Henne, S. and Prévôt, A. (2011). Influence of meteorology on PM 10 trends and variability in Switzerland from 1991 to 2008. *Atmospheric Chemistry and Physics* **11** 1813–1835.
9. Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association* **80** 580–598. MR0803258 MR0803258
10. Breitung, J. and Roling, C. (2015). Forecasting inflation rates using daily data: A nonparametric MIDAS approach. *Journal of Forecasting* **34** 588–603. MR3416308 MR3416308
11. Chakraborty, T., Chattopadhyay, S. and Ghosh, I. (2019). Forecasting dengue epidemics using a hybrid methodology. *Physica A: Statistical Mechanics and its Applications* **527** 121266.
12. Clements, M. P. and Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business & Economic Statistics* **26** 546–554. MR2459351 MR2459351
13. Clements, M. P. and Galvão, A. B. (2009). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics* **24** 1187–1206. MR2750284 MR2750284
14. Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics* **20** 134–144. MR1940633 MR1940633
15. Engle, R. F., Ghysels, E. and Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics* **95** 776–797.
16. Feng, S., Gao, D., Liao, F., Zhou, F. and Wang, X. (2016). The health effects of ambient PM2.5 and potential mechanisms. *Ecotoxicology and Environmental Safety* **128** 67–74.
17. Foroni, C., Marcellino, M. and Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society Series A: Statistics in Society* **178** 57–82. MR3291761 MR3291761
18. Ghysels, E., Kvedaras, V. and Zemlys, V. (2020). Mixed data sampling (MIDAS) regression models. In *Handbook of Statistics*, **42** 117–153. Elsevier.
19. Ghysels, E. and Marcellino, M. (2018). *Applied economic forecasting using time series methods*. Oxford University Press.
20. Ghysels, E., Santa-Clara, P. and Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regression models. *CIRANO Working Papers*.
21. Ghysels, E., Santa-Clara, P. and Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics* **131** 59–95. MR2288650 MR2288650
22. Ghysels, E., Sinko, A. and Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews* **26** 53–90. MR2339264 MR2339264
23. Giannone, D., Reichlin, L. and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics* **55** 665–676.
24. Gilmore, C. G. (2001). An examination of nonlinear dependence in exchange rates, using recent methods from chaos theory. *Global Finance Journal* **12** 139–151.
25. Gustin, M., McLeod, R. S. and Lomas, K. J. (2019). Can semi-parametric additive models outperform linear models, when forecasting indoor temperatures in

free-running buildings? *Energy and Buildings* **193** 250–266.
26. Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models* **43**. CRC press. MR1082147 MR1082147
27. Kim, S.-J., Koh, K., Boyd, S. and Gorinevsky, D. (2009). \ell_1 trend filtering. *SIAM Review* **51** 339–360. MR2505584 MR2505584
28. Liew, V. K.-S. (2004). Which lag length selection criteria should we employ? *Economics Bulletin* **3** 1–9.
29. Mogliani, M. and Simoni, A. (2021). Bayesian MIDAS penalized regressions: estimation, selection, and prediction. *Journal of Econometrics* **222** 833–860. MR4236449 MR4236449
30. Peña, M., Vázquez-Patiño, A., Zhiña, D., Montenegro, M. and Avilés, A. (2020). Improved rainfall prediction through nonlinear autoregressive network with exogenous variables: a case study in Andes high mountain region. *Advances in Meteorology* **2020** 1828319.
31. Qi, M. (1999). Nonlinear predictability of stock returns using financial and economic variables. *Journal of Business & Economic Statistics* **17** 419–429.
32. Ramdas, A. and Tibshirani, R. J. (2016). Fast and flexible ADMM algorithms for trend filtering. *Journal of Computational and Graphical Statistics* **25** 839–858. MR3533641 MR3533641
33. Sadhanala, V. and Tibshirani, R. J. (2019). Additive models with trend filtering. *Annals of Statistics* **47** 3032–3068. MR4025734 MR4025734
34. Simon, T., Fabsic, P., Mayr, G. J., Umlauf, N. and Zeileis, A. (2018). Probabilistic forecasting of thunderstorms in the Eastern Alps. *Monthly Weather Review* **146** 2999–3009.
35. Tangang, F. T., Tang, B., Monahan, A. H. and Hsieh, W. W. (1998). Forecasting ENSO events: A neural network–extended EOF approach. *Journal of Climate* **11** 29–41.
36. Tibshirani, R. J. (2013). The lasso problem and uniqueness. *Electronic Journal of Statistics* **7** 1456–1490. MR3066375 MR3066375
37. Tibshirani, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Annals of Statistics* **42** 285–323. MR3189487 MR3189487
38. Tibshirani, R. J. and Taylor, J. (2011). The solution path of the generalized lasso. *Annals of Statistics* **39** 1335–1371. MR2850205 MR2850205
39. Tseng, P. (2001). Convergence of a block coordinate descent method for non-differentiable minimization. *Journal of Optimization Theory and Applications* **109** 475–494. MR1835069 MR1835069
40. Turner, P. R. (1994). *Numerical Analysis*. Macmillan Education UK. https://doi.org/10.1007/978-1-349-13108-2
41. Wang, K., Deng, C., Li, J., Zhang, Y., Li, X. and Wu, M. (2017). Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network. *Epidemiology & Infection* **145** 1118–1129.
42. Wang, Y.-X., Smola, A. and Tibshirani, R. (2014). The falling factorial basis and its statistical applications. In *International Conference on Machine Learning* 730–738. PMLR.
43. Wei, C.-C. (2012). Wavelet support vector machines for forecasting precipitation in tropical cyclones: Comparisons with GSVM, regression, and MM5. *Weather and Forecasting* **27** 438–450.
44. Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC press. MR3726911 MR3726911
45. Xu, Q., Chen, L., Jiang, C. and Liu, Y. (2022). Forecasting expected shortfall and value at risk with a joint elicitable mixed data sampling model. *Journal of*

Forecasting **41** 407–421. MR4396879 MR4396879

46. Yang, Z., Yang, J., Li, M., Chen, J. and Ou, C. (2021). Nonlinear and lagged meteorological effects on daily levels of ambient PM2.5 and O3: Evidence from 284 Chinese cities. *Journal of Cleaner Production* **278** 123931.
47. Zhang, J. J., Wei, Y. and Fang, Z. (2019). Ozone pollution: a major health hazard worldwide. *Frontiers in Immunology* **10** 2518.

Note: This list reflects references listed in the original paper as accurately as possible with no attempt to correct errors.