Extensions on the Fractional Differencing Methodology for Portfolio Construction

Ernest Digore

In this paper, we investigate the performance of a momentum trading strategy augmented by an ARFIMA(p,d,q) statistical model. By incorporating long-memory features into the modelling of stock returns, we aimed to achieve more accurate predictions and subsequently, superior trading outcomes. Our research was motivated by the quest to further refine trading strategies and increase their profitability while accounting for market complexities. The main focus is on the thorough dissection of the Sowell (1992) Maximum Likelihood Estimation methodology of ARFIMA parameters based on the pioneering results in Hosking (1981) and recommended by Dahlhaus (1988). We pinpoint the main advantages and disadvantages of this methodology, suggested improvements, and describe in details the application process towards a momentum trading strategy that extends the work of Chitsiripanich et al. (2022).

The general ARFIMA(p,d,q) can be written in a compact form as:

$$\Phi(L)(1-L)^d z_t = \Theta(L)\varepsilon_t \tag{1}$$

and we introduce the Normal Likelihood function that is used to estimate the ARFIMA parameters, namely the ARMA coefficients and the fractional difference parameter d:

$$f\left(Z_T, \mathbf{\Sigma}\right) = (2\pi)^{-T/2}|\mathbf{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2} Z_T' \mathbf{\Sigma}^{-1} Z_T\right\} \tag{2}$$

where Z_T is a sample of T observations normally distributed with $\mu = 0$ and covariance matrix Σ, which is of a Toeplitz form. The complexity around the computation of the covariance matrix revolves around implementing ten distinctive functions in the style of Sowell (1992) and Chung (1994) that involve multiple series of hypergeometric functions and gamma functions.

Lastly, we perform an unconstrained maximum log-likelihood estimation of the ARFIMA parameters based on the Normal Likelihood function:

$$\hat{\nu} = \underset{\nu\in\Omega}{\arg\max} f_l\left(Z_T, \Sigma(T,\nu)\right) \quad \text{with} \quad \hat{\nu} = [\hat{d}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}] \tag{3}$$

where f_l is the logarithm of the Normal Likelihood function, Ω is the parameter space, a finite-dimensional subset of Euclidean space, $\hat{d}$ is the fractional difference estimated parameter, $\hat{\boldsymbol{\phi}}$ - a vector of AR(p) estimated coefficients, and $\hat{\boldsymbol{\theta}}$ - a vector of estimated MA(q) coefficients.

Due to the fact that the likelihood equations $\frac{\partial f_l(Z_T,\Sigma(\nu)}{\partial \nu} = 0$ cannot be solved explicitly for an estimator $\hat{\nu} = \hat{\nu}(Z_T)$, we introduce an iterative process based on an updating formula with an initial guess ($\hat{\nu_1}$) as recommended in Sowell (1992). This is a two-step implementation: choosing a grid of d parameters (from -0.49 to 0.49 to follow the restriction of ARFIMA), fractionally differencing the series, and estimating the ARMA parameters, based on that series. The initial guess will be the parameters that portray the lowest variance of the innovations from this procedure.

In the style of Doornik and Ooms (2004), the best linear prediction of $z_{T+H|T}$, given the information in Z_T and knowing the estimated parameters $\hat{\nu} = [\hat{d}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}]$ of the ARFIMA process, is given by:

$$\widehat{z}_{T+H|T} = (\gamma_{T-1+H} \cdots \gamma_H) \left(\boldsymbol{\Sigma}_T\right)^{-1} Z_T \tag{4}$$

$\left(\boldsymbol{\Sigma}_T\right)^{-1} Z_T$ could easily be computed via a Durbin-Levinson algorithm (see Durbin (1960)), which is used for the inversion of finite Toeplitz matrices. Essentially, computing $\gamma(0), ..., \gamma(T+H)$ for the input parameters $\hat{\nu}$ allows the forecasting of future H data points of the time series Z_T.

As echoed by Chitsiripanich et al. (2022), applying the first difference transformation to log-returns might constitute an overly drastic modification, leading to an over-differenced and potentially less informative dataset. This concern is particularly relevant in the context of ARFIMA(0,d,0) models, the forecasting performance of which has been critically evaluated by Ellis and Wilson (2004). However, it is essential to note that these critiques were primarily levelled at analyses of conventional stock returns, which are generally not believed to possess substantial long-term memory. To solve the conundrum of having a stationary (quasi-stationary) input for the ARFIMA model while not eliminating the entire memory or predictive power from the data, it would be mindful to apply a two-step approach by first obtaining a time series $(1-L)^{d_1} z_t$ and then estimating d_2 in the classical sense of ARFIMA.

For our analysis, we choose two values of d_1: 0.4 and 0.9. First ensures with a high probability that the transformed time series are stationary and allows the preservation of memory to the maximum extent possible, while the latter still preserves memory in the data, but allows the evaluation of the autocovariance function without implying roots close to the unit circle. Consequently, we choose an $ARFIMA(2, d_1 + d_2, 2)$, parsimonious model as a compromise between computation time and estimation power.

The trading strategy, per se, consists of forecasting future values of a stock's universe at each point in time for a rolling window of 250 days, ranking the forecasts, creating quantile buckets, and going long the top bucket and short the bottom bucket. We chose a value-weighted strategy with weekly rebalancing, a period of backtesting of 5 years - December 2015 to December 2020, and a randomised set of paths that involve having a portfolio of 30 stocks at each point in time. The same data set as in Chitsiripanich et al. (2022) is used; namely, the Asness et al. (2013) universe and for the purpose of comparison we showcase the strategies $ARFIMA(2, 0.4+d_2, 2)$, $ARFIMA(2, 0.9+d_2, 2)$, $FI(0.9)$ as in Chitsiripanich et al. (2022) but with the same stock universe, and S&P500 as a proxy for the market. The bucket size is chosen to be $q = 0.2$, as it was proved to be the best performing one.

Metric	$d_1 = 0.4$	$d_1 = 0.9$	$FI(0.9)$	S&P500
Annual Return	6.56%	-2.61%	-3.38%	14.96%
Volatility	11.38%	14.57%	25.62%	19.30%
Sharpe Ratio	0.58	-0.18	-0.13	0.29
Sortino Ratio	0.83	-0.24	-0.17	0.60
STARR	0.0173	-0.0028	-9e-05	0.0201
Skewness	-0.162	-0.207	-1.410	-1.144
Kurtosis	7.571	7.556	22.32	25.303
Maximum Drawdown	-17.39%	-50.10%	-58.51%	-67.74%
Average Drawdown	-4.53%	-19.10%	-14.81%	-7.33%
Drawdown Duration	218 days	1035 days	666 days	145 days
Beta	-0.0186	-0.0096	-0.0687	1.0000
Hit Ratio	53.40%	49.88%	52.36%	56.35%
VaR_{95}	1.05%	1.46%	2.10%	1.75%
CVaR_{95}	1.61%	2.23%	4.10%	3.13%
Turnover	79.6%	79.5%	78.5%	NA

Table 1: Performance metrics for different strategies and the market index.

$ARFIMA(2, 0.4 + d_2, 2)$ outperforms all the other strategies and the market on a risk-adjusted returns basis. This is entailed from the higher Sharpe and Sortino ratios, lower Skewness and Kurtosis, much lower Drawdowns, and lower risk-measures such as VaR_{95} & CVaR_{95}. Two other key elements of these returns are the absence of correlation with the market and even lower crash risk, the dynamics of which were captured during the Covid-19-related crisis. One also has to be mindful of the implied high turnover.

However, these findings should be considered cautiously, given the limitations of the data sample scope. Therefore, the main disadvantage of the results is the small sample of stocks used for the generation of the quantile buckets due to computational constraints pertaining to the EMLE of ARFIMA parameters. Nevertheless, through an inference process, we expect the results to be even more attractive with greater sample size. Future research could leverage more substantial computing resources, extend the stock selection, or apply alternate estimation methodologies such as indirect estimation of ARFIMA as in Martin and Wilkins (1999) or other parametric and semiparametric methodologies recommended by Fox and Taqqu (1986) and tested by Reisen et al. (2001).

References

C. S. Asness, T. J. MOSKOWITZ, and L. H. PEDERSEN. Value and momentum everywhere. *The Journal of Finance*, 68(3):929–985, 2013. ISSN 00221082, 15406261. URL http://www.jstor.org/stable/42002613.

S. Chitsiripanich, M. S. Paolella, P. Polak, and P. S. Walker. Momentum without crashes. Nov 2022. Swiss Finance Institute Research Paper No. 22-87, Available at SSRN: https://ssrn.com/abstract=4280465 or http://dx.doi.org/10.2139/ssrn.4280465.

C.-F. Chung. A note on calculating the autocovariances of the fractionally integrated arma models. *Economics Letters*, 45(3):293–297, 1994. ISSN 0165-1765. doi: https://doi.org/10.1016/0165-1765(94)90026-4. URL https://www.sciencedirect.com/science/article/pii/0165176594900264.

R. Dahlhaus. Small Sample Effects in Time Series Analysis: A New Asymptotic Theory and a New Estimate. *The Annals of Statistics*, 16(2):808 – 841, 1988. doi: 10.1214/aos/1176350838. URL https://doi.org/10.1214/aos/1176350838.

J. A. Doornik and M. Ooms. Inference and forecasting for arfima models with an application to us and uk inflation. *Studies in Nonlinear Dynamics Econometrics*, 8(2), 2004. doi: doi:10.2202/1558-3708.1218. URL https://doi.org/10.2202/1558-3708.1218.

J. Durbin. The fitting of time-series models. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, 28(3):233–244, 1960. ISSN 03731138. URL http://www.jstor.org/stable/1401322.

C. Ellis and P. Wilson. Another look at the forecast performance of ARFIMA models. *International Review of Financial Analysis*, 13(1):63–81, 2004. URL https://ideas.repec.org/a/eee/finana/v13y2004i1p63-81.html.

R. Fox and M. S. Taqqu. Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, 14(2):517–532, 1986. ISSN 00905364. URL http://www.jstor.org/stable/2241233.

J. R. M. Hosking. Fractional differencing. *Biometrika*, 68(1):165–176, 1981. ISSN 00063444. URL http://www.jstor.org/stable/2335817.

V. L. Martin and N. P. Wilkins. Indirect estimation of arfima and varfima models. *Journal of Econometrics*, 93(1):149–175, 1999. ISSN 0304-4076. doi: https://doi.org/10.1016/S0304-4076(99)00007-X. URL https://www.sciencedirect.com/science/article/pii/S030440769900007X.

V. Reisen, B. Abraham, and S. Lopes. Estimation of parameters in arfima processes: A simulation study. *Communications in Statistics - Simulation* *and Computation*, 30(4):787–803, 2001. doi: 10.1081/SAC-100107781. URL https://www.tandfonline.com/doi/abs/10.1081/SAC-100107781.

F. Sowell. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53(1-3):165–188, 1992. URL https://EconPapers.repec.org/RePEc:eee:econom:v:53:y:1992:i:1-3:p:165-188.