High-Dimensional Temporal Disaggregation

Luke Mosley

STOR-i Centre for Doctoral Training, Department of Mathematics and Statistics, Lancaster University, United Kingdom

Idris Eckley

Department of Mathematics and Statistics, Lancaster University, United Kingdom

Alex Gibberd†

Department of Mathematics and Statistics, Lancaster University, United Kingdom

E-mail: a.gibberd@lancaster.ac.uk

Summary. Temporal disaggregation is a method commonly used in official statistics to enable high-frequency estimates of key economic indicators, such as GDP. Traditionally, such methods have relied on only a couple of high-frequency indicator series to produce estimates. However, the prevalence of large, and increasing, volumes of administrative and alternative data-sources motivates the need for such methods to be adapted for high-dimensional settings. In this work, we propose a novel sparse temporal-disaggregation procedure and contrast this with the classical Chow-Lin method. We demonstrate the performance of our proposed method through simulation study, highlighting various advantages realised. We also explore its application to disaggregation of UK gross domestic product data, demonstrating the method's ability to operate when the number of potential indicators is greater than the number of low-frequency observations.

Keywords: Temporal Disaggregation; Official Statistics; Fast Indicators; Model Selection; High-Dimensional Statistics; Generalised Least Squares

†*Address for correspondence: Department of Mathematics and Statistics, Lancaster University, Bailrigg, LA1 4YF*

1. Motivation

Our work seeks to address an important question in economic statistics, namely is there a principled approach by which one can provide more frequent and interpretable estimates of headline economic variables using existing (more infrequent) measures and supplemental data. Specifically, our work addresses the task of temporal disaggregation, constructing high frequency estimates of low frequency time series, and asks how this task can be effectively tackled in a high-dimensional setting. Disaggregating to the high frequency is usually done by incorporating a few high frequency indicator series into a regression model that are believed to model the short-term dynamics of the variable of interest. However, the prevalence of large, and increasing, volumes of administrative and alternative data-sources motivates the need for such methods to be adapted for high-dimensional settings. These are settings where we want to include more indicator series into the model than the number of data points observed for the low frequency series. In such settings, traditional temporal disaggregation methods become statistically incapable of providing estimates. By integrating methods from the high-dimensional statistics literature, our key contribution is to create a regularised M-estimation framework for well-established Chow-Lin temporal disaggregation procedure (Chow and Lin, 1971), which can build robust and interpretable estimates in high-dimensional scenarios.

This work is motivated by the challenging task of performing high frequency disaggregation for UK national GDP, moving from a quarterly to a monthly resolution. For this task, there are a considerable number of indicator series that one may wish to use. In our application we consider survey based series, such as the monthly business survey (MBS) in both services and production, alongside VAT data, retail sales indices, and several novel indicators such as traffic flows at ports and on roads. In total, we consider 97 indicators, all of which are collected at a monthly frequency. Given the significant interest in fast measurements of economic activity, the UK's Office for National Statistics (ONS) has developed a monthly GDP index, and published this statistic since May 2018. Even though a monthly index exists in this case, there is still great interest in performing temporal disaggregation, the reasons are threefold. Firstly, the monthly index is an output based measure, however economists may also be interested in both expenditure and income based estimates. Since, temporal disaggregation can be applied to any output stream, either expenditure or income based measures could be used. The resulting high-frequency estimate can thus compliment the existing output based index. Secondly, due to the construction of the index, publication lags the period of measurement (an issue common to most economic statistics). However, temporal disaggregation can potentially be used to find indicators that are relevant and updated more frequently providing a faster estimate of the output statistic. National Statistics Institutes (NSIs) are actively developing so-called fast-indicators for exactly this purpose and in this article we consider several of these in the form of traffic data. Finally, one of the key issues surrounding the fast release of data is in understanding the associated short-term movements. To this end, temporal disaggregation using interpretable indicator series can provide insight by highlighting which indicators are driving movement.

Through extensive simulation studies we investigate the performance of our approach in estimating high frequency disaggregated series in both standard and high-dimensional scenarios. We also compare against the established Chow-Lin method in standard dimensional settings. In the quarterly-to-monthly GDP disaggregation application we demonstrate that our estimated model not only aligns with economic intuition, but also achieves better tracking of the published monthly GDP index when compared against the Chow-Lin method.

2. Temporal Disaggregation

The temporal disaggregation problem is as follows. Given the $(n \times 1)$ vector y_l of low frequency observations of an economic aggregate and given the $(m \times p)$ matrix X_h of high frequency observations on p related indicators, we wish to estimate the $(m \times 1)$ vector y_h of high frequency unobserved observations. For example, using quarterly imports and exports series to disaggregate annual trade data would mean n represents the number of observed years, $m = 4n$ the number of quarters and $p = 2$ the number of indicator series available. We can set up the linear regression at the high frequency as $y_h = X_h\beta + u_h$, where β is the $(p \times 1)$ vector of unknown parameters and u_h is a vector of $(m \times 1)$ random disturbances such that $u_h \sim N(0, V_h)$. As the dependent variable y_h is not observed, the procedure is to temporally aggregate the regression to the observed low frequency via the aggregation matrix $C = I_n \otimes 1_k$, where $\otimes$ is the Kronecker product and 1_k is a k-dimensional row vector of ones where k is the number of high frequency periods between each low frequency observation. This is assuming the data is flow data as the aggregation will be a sum. For index data ones would be replaced by $1/k$ as we average. For stock data aggregation would be made by interpolation where we impute the first or last high frequency period.

We then have a linear regression at the observed low frequency: $y_l = X_l\beta + u_l$, where $y_l = Cy_h$, $X_l = CX_h$ and $u_l = Cu_h$ with $u_l \sim N(0, V_l)$ for $V_l = CV_hC^T$. The best linear unbiased estimator of y_h consistent with the aggregation constraint $y_l = Cy_h$ is given by

$$\hat{y}_h = X_h\hat{\beta} + V_hC^TV_l^{-1}(y_l - X_l\hat{\beta})\,, \tag{1}$$

where $\hat{\beta}$ is the Generalised Least Squares (GLS) estimate of β in the low frequency observed model given by

$$\hat{\beta} = (X_l^TV_l^{-1}X_l)^{-1}X_l^TV_l^{-1}y_l\,. \tag{2}$$

Clearly, the covariance matrix V_h is assumed to be known to construct these estimates and hence in practical problems we must estimate it. Chow and Lin (1971) propose assuming a known structure of V_h by assuming the disturbance series u_h follows an AR(1) model. I.e.

$$\begin{aligned} u_{h,t} &= \rho u_{h,t-1} + \epsilon_{h,t}\,, && (3)\\ |\rho| &< 1\,, && (4)\\ \epsilon_{h,t} &\sim N(0,\sigma^2)\,. \end{aligned}$$

The estimation of the unknown parameters (β, ρ, σ^2) are found by means of maximising the log-likelihood function of the implied low frequency model $\ell(\beta, \sigma^2|\rho)$ by optimising via a grid search over the stationary $(-1, 1)$ domain of ρ. The assumed structure in

(3) is desirable as if y_h and X_h were both non-stationary, then under (4), they are co-integrated in the sense of Engle and Granger (1987). A very attractive property when forecasting time series. Other authors (Fernandez, 1981; Litterman, 1983) propose non-stationary processes in (3) which generally result in smoother high frequency estimates at the expense of forecasting performance.

3. Sparse Temporal Disaggregation

Despite the popularity of Chow and Lin (1971) to compile national accounts across Europe (Eurostat, 2018), the method runs into several shortcomings when operating in data-rich environments NSIs now find themselves in. In moderate and high dimensions, the behaviour of the Chow-Lin procedure faces several statistical challenges: a) excessive variance in $\hat{\beta}$ impacts interpretation as all indicator series are included in the model by default; b) inconsistent estimation of the AR(1) parameter ρ leads to poor performance in estimating the high frequency series; c) unreliable estimation of σ^2 leads to uncertainty in estimating β; d) lack of interpretation into which indicator series are most relevant. Furthermore, when $p > n$, the matrix $X_l^T V_l^{-1} X_l$ needing to be inverted in (2) becomes rank-deficient and thus a unique inverse no longer exists, and so a disaggregated estimate cannot be found.

To resolve the aforementioned shortcomings of current temporal disaggregation methods, we provide a general regularised M-estimation framework that allows us to encompass a variety of penalty functions in the Chow-Lin regression framework to accomplish temporal disaggregation in moderate and high dimensions. Specifically, we propose to study estimators of the form:

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \left\{ \underbrace{\left\| V_l^{-1/2}(y_l - X_l\beta) \right\|_2^2}_{\text{Chow-Lin Cost Function}} + \underbrace{P_\lambda(\beta)}_{\text{Regulariser}} \right\}. \tag{5}$$

This estimator incorporates a *regularising penalty function* in conjunction with the classic Chow-Lin cost function to encode the assumption of sparsity. It does this by shrinking coefficients of indicator series, β, towards zero that cause a large least squares score in the Chow-Lin cost function. By doing so, we simultaneously select important indicator series and estimate their regression coefficients with sparse estimates. This will significantly reduce the variance in moderate dimensions and enable accurate estimators in high dimensions.

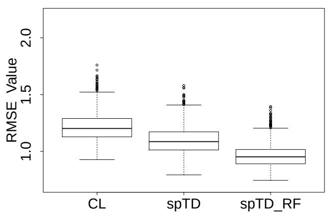
The regulariser function, $P_\lambda(\beta)$, is indexed by the regularisation parameter $\lambda \geq 0$ that controls the degree of shrinkage. This function may take various forms depending on the assumptions required by the user, see Bühlmann and Van De Geer (2011) for options. We study the the LASSO penalty by Tibshirani (1996) in our work, i.e. $P_\lambda(\beta) = \lambda||\beta||_1$, due to its desirable property of yielding sparse estimates while maintaining convexity. The algorithm we propose is as follows:

1. Temporally aggregate the indicators via C and perform a GLS rotation of the data using an initial value $\rho \in (-1, 1)$. I.e. we are working with $y = V_l^{-1/2} y_l$ and $X = V_l^{-1/2} X_l$.

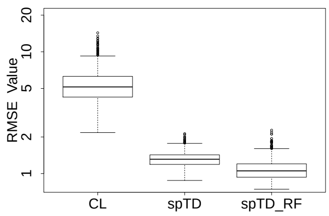
2. Use the Least Angle Regression (LAR) algorithm by Efron et al. (2004) to compute full piecewise linear solution paths $\hat{\beta}_\lambda$ for a range of λ.
3. Re-fit the selected sparse support from $\hat{\beta}_\lambda$ back into least squares to de-bias the LASSO estimates.
4. Tune the model for λ using Bayesian Information Criterion (BIC) (Schwarz, 1978) using degrees of freedom equal to $K_\lambda = |\{r : \hat{\beta}(\lambda)_r \neq 0\}|$ and variance estimator $\hat{\sigma}^2 = ||V_l^{-1/2}(y_l - X_l\hat{\beta}(\lambda))||_2^2/2(n - K_\lambda)$.
5. Optimise over $\rho \in (-1, 1)$ and use $\hat{\rho}$

4. Simulation Study and GDP Data Application

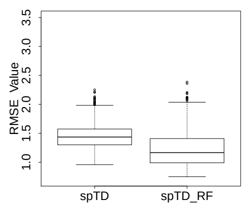
We perform an extensive simulation study of several parameter scenarios and deliver accurate estimates in the high dimensional setting and a great improvement on Chow-Lin in standard dimensional settings. We consider annual-to-quarterly disaggregation using $n = 100$ years, with $p = 30$ or 90 for standard dimensions and $p = 150$ for high dimensions. We consider $\rho = 0.2, 0.5$ and 0.8 for low, medium and high auto-correlation present in the residuals and also consider both stationary and non-stationary time series for y_h and X_h. Figure 1 (a) and (b) show the improved performance of our approach both with and without the re-fit step (in step 3 of algorithm) on Chow-Lin in standard dimensions using $p = 30$ and 90 respectively in estimating y_h. While Figure 1 (c) shows our accurate performance when $p = 150$.



(a)



(b)



(c)

Fig. 1: Boxplots comparing RMSE values of $\hat{y_h}$ for CL, spTD and spTD_RF. Where a, b and c represent $p = 30$, 90 and 150 respectively.

As described in Section 1, we attempt to perform a quarterly-to-monthly disaggregation of UK national GDP and assess performance against the published monthly GDP index developed by the ONS. Using data from 2008 Q1 to 2020 Q2 ($n = 50$ quarters), we compare our performance against Chow-Lin using 10 monthly indicator series and assess how well we do using 97 monthly indicator series; the high-dimensional setting. We obtain a lower RMSE than Chow-Lin (985.13 compared to 1055.74) when using 10 indicator series in estimating monthly GDP and even greater performance in the high-dimensional setting (RMSE = 749.63) showing the advantage of working in high-dimensions. Not only do we get more accurate estimates, our method informs us on the most relevant indicator series used to derive the estimates; informing us on the main driving forces behind monthly GDP.

References

Bühlmann, P. and Van De Geer, S. (2011) *Statistics for high-dimensional data: methods, theory and applications.* Springer Science & Business Media.

Chow, G. C. and Lin, A.-l. (1971) Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The review of Economics and Statistics*, 372–375.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R. et al. (2004) Least angle regression. *The Annals of statistics*, **32**, 407–499.

Engle, R. F. and Granger, C. W. (1987) Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251–276.

Eurostat (2018) ESS guidelines on temporal disaggregation, benchmarking and reconciliation.

Fernandez, R. B. (1981) A methodological note on the estimation of time series. *The Review of Economics and Statistics*, **63**, 471–476.

Litterman, R. B. (1983) A random walk, markov model for the distribution of time series. *Journal of Business & Economic Statistics*, **1**, 169–173.

Schwarz, G. (1978) Estimating the dimension of a model. *The annals of statistics*, **6**, 461–464.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**, 267–288.