Least Squares-Optimal Transport (LS-OT) Regression for varying time delays

B. Tsakam-Sotché, brice.tsakam@gmail.com
Horgen, February 2023

Abstract

While analysing time series data, conventional linear regression methods leave us the choice with one of two assumptions: either the dependent and independent variables are synchronous or fixed time lags are assumed.
However the time lag might be varying over time, or noisy. The time lag becomes an additional source of error beside the usual (measurement) noise. Ignoring the variable time lag amounts to a specification error. We propose to use optimal transport cost as a way to account for the time lag uncertainty and make ordinary least squares (OLS) robust to this type of specification error. This enhancement to the celebrated OLS regression model of noise easily generalises to multivariate regression, and alternative models such as Partial Least Square.

Introduction

We investigate the simple regression for non-synchronous time series. This type of data is common in practice in a variety of contexts. We propose a method that is relevant when all the usual assumptions for OLS are fulfilled but there is still an incorrect model specification in the sense that the time lag variation is significant in comparison to the variation of the dependent variable. This applies to situations where the dependent and independent variables are connected with a transmission link that is not instantaneous and might even exhibit shocks in the transmission delay.
Current methods for this situation include OLS with multiple time lags, but with time lag averaging to zero, the best regressor is one that assumes no time lag at all. Hence we compare the proposed method with simple OLS regression with no time lag assumption as a baseline.
We review OLS regression, Optimal Transport and the proposed Least-Square Optimal-Transport (LS-OT) regression in Section 1. Section 2 presents the comparison of OLS and LS-OT using synthetic data. We conclude with a discussion of early results, limitations and further research.

Section 1

1.1 OLS regression

We first recall OLS basic assumptions:

- The dependent variable is a linear function of the independent variables: y = Beta * X
- The observations (xi,yi) are independent and identically distributed (iid)
- The non-dependent variables are linearly independent
- residuals eta = y-y_est are normal, uncorrelated with fixed variance sigma

The OLS [2] method aims at inferring the dependent variable yt from the contemporary independent variable xt and time lags thereof xt-1, xt-2, etc. The linear dependency factor beta, which is identified through an optimization process that minimises OLS_{cost} defined as follows:
OLS_{cost} = Sum{(y_t - Beta * T * X_t)^2}
Regression models are evaluated using the coefficient of determination R2 which is related to the ratio of unexplained over total variance:
Rsquared = 1 - Sum{(y - y_est)^2} / Sum{(y - E(y))^2}, Rsquared = 1 in case of a perfect fit and 0 for a poor fit.

1.2 Optimal Transport

Let's consider the Optimal Transport cost [1]:
OT_{cost} = Sum{x,y C(x,T(x)) ∆M(x)} where T^*(M)=V
Where *M* and *V* are measures and T^*(*M*) is the optimal transportation plan that moves the mass of *M* into *V* at a minimal cost given the Cost matrix C.

In comparison, the OLS uses the Euclidean distance as cost and sets T to be a linear function parameterized with Beta: T(x) = Beta * T * X.
We observe that this is identical to the Wasserstein distance when the transport cost is defined by the Euclidean distance except that no error is allowed in the time axis which can be seen as an infinite cost.

1.3 Least-Squares Optimal-Transport regression

Now in the OT distance minimization framework, we have the flexibility to allow a time lag and handle data that is not perfectly synchronous. We state the regression problem as an optimization problem combining the OLS and OT costs, with a trade-off parameter lambda:

Argmin Beta: {OLS_{cost}(Beta) + lambda * OT_{cost}(Beta)}
parameterized with an isotropic cost matrix

Section 2

2.1 Synthetic data generator

In the case of OLS the input data is assumed to be synchronous. The time lag is assumed to be fixed. We now consider a data generation process with explicit random time lag:
y_t = beta * X(t+eta_t) + eta
X_t is a vector: [x_t, 1], beta is a constant vector: [b, alpha]
Eta is normal with variance s_e.
The time lag noise eta_t is normal with variance s_e_t
Equipped with this data generator, we proceed to estimate beta with different methods and compare results.

2.2 Implementation details

We use python jax [3] automatic differentiation software for the OT cost gradients. The chain rule allows to backpropagate the gradient to regression model parameters, then gradient descent converges to the desired minima.
We compensate for the bias in the OT distance calculation as in [4].
We run 100 simulations and report the average and variance of beta using the OLS and the Transport regression. We allow only positive time lags (ie positive time lag noise eta_t), which is symmetrical to the case with negative time lags.
(The complete jupyter notebook for data generation and regression experiments presented here is available at: [https://colab.research.google.com/drive/1MH_vGrtf-cuPKtxBQbokLsDRRrXr5ero])

4. Results & Discussion

4.1 Results

As shown in Figure 1, the proposed method consistently improves on OLS on the regression task given time varying lag or the order of magnitude of the time between samples. We observe that the improvement is consistent across samples.

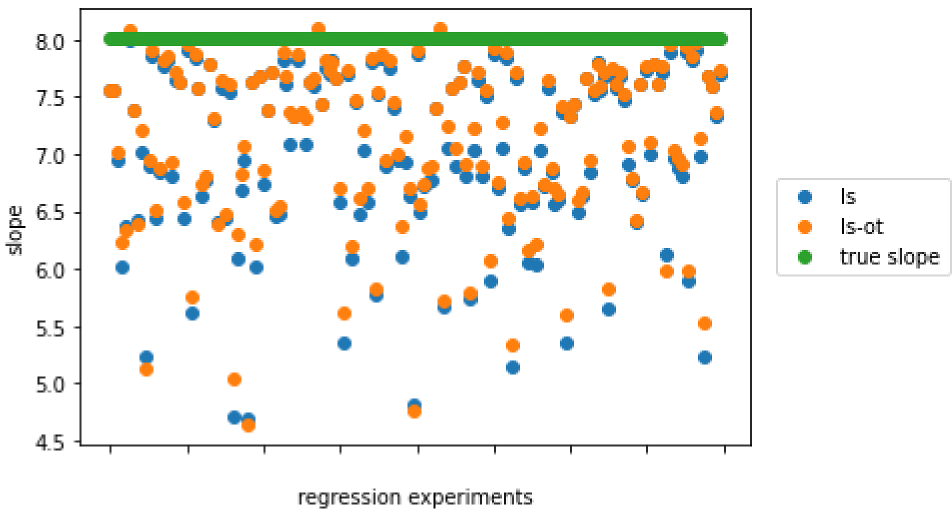

Figure 1: ls shows OLS estimated Beta, LS-OT estimated and the true slope is for the true Beta (as generator parameter)

The LS-OT method is a favourable option when non-synchronous data is at hand and the preferred model when expert knowledge on the time lag variance is available. This method can easily be extended to partial least square, autoregressive models that exhibit variable time lags and vector variants thereof.

4.2 Further research

This method is severely limited by the OT accuracy and numerical stability. We could only use very short time series with 10-20 samples because of the limited accuracy of the gradient descent OT solver. The LS-OT numerical issues forced us to manually test the lambda parameter to identify a stable range (1e-6 to 1e-4). Solving these two issues would make LS-OT approach a practical method to extend OLS to cases where the time lag induced noise is significant (eta_t is comparable to eta).

In addition, we may apply the singular value decomposition (SVD) on the transportation plan to get the main (largest) eigenvalues and corresponding eigenvectors as a rough characterisation of the time lag. A single mode (of fact decaying eigenvalues) SVD would suggest that it is better to revert to OLS with the appropriate fixed time lag while multiple significant eigenvalues means that the time lag is arguably not fixed and diverse.

References

[1] Wikipedia article for Optimal Transport. Retrieved in February 2023 https://en.wikipedia.org/wiki/Transportation_theory_(mathematics)#Abstract_formulation_of_the_problem

[2] Wikipedia article for Ordinary Least Squares. Retrieved in February 2023 https://en.wikipedia.org/wiki/Ordinary_least_squares

[3] Github repo for jax (February 2023) https://github.com/google/jax

[4] Hicham Janati, Marco Cuturi, Alexandre Gramfort, “Debiased Sinkhorn barycenters”, retrieved Feb 2023 https://arxiv.org/abs/2006.02575