Rolling window experiments

New Code files:
run_arima_ets_rolling_origin.py -> runs the base arima and ets rolling window forecasts
run_ml_method_rolling_origin.py -> runs the ML reconciliation for the rolling windows
run_hts_benchmarks_rolling_origin.py -> runs the reconciliation benchmarks for the rolling window
calculate_rolling_origin_errors.py -> this code will read all the base and reconciled forecasts and output a csv file with the percentage relative improvement of each reconciliation method [the order of the columns shows order of the method performance]

The forecasts are only for one-step ahead

- Each rolling window has a set of fitted values, actual values corresponding to the fitted data, forecasts 1-step and the actual values for the 1-step ahead forecasts

- For all reconciliation approaches the reconciliation approach is trained/ parameters are found after every 10th rolling window (we discussed this approach to reduce the time complexity for re-estimating parameters for each rolling-window)
    - For example, if we have 20 rolling windows for 1-10 rolling windows the parameters are found for the reconciliation method using the 1st windows fitted values. And the forecasts are reconciled using these parameters for all 1-10 rolling windows. Then the parameters are re-estimated from the 11th rolling window and these parameters used to reconcile the forecasts across 11-20th rolling windows.

For ML methods:
Case 1 -> Only the bottom level error as the validation loss
Case 2 -> Whole hierarchy is considered in the validation loss

Prison

ARIMA

	ols	mintsample	wls	bottomup	mintshrink	case1_lambda_[0.01, 0.09]
Australia	5.82	-21.17	-24.63	-129.82	-750.49	-22500.18
State	2.73	-2.50	-2.83	-46.74	-622.88	-6072.20
Gender	4.63	3.12	4.26	-35.98	-614.71	-5520.12
Legal	13.45	11.18	11.63	-4.37	-561.59	-2405.30
Indigenous	11.29	13.12	13.43	0.00	-545.22	-1986.76
Overall	7.60	-0.86	-1.49	-49.66	-628.59	-8795.74

ETS

	mintsample	ols	wls	bottomup	mintshrink	case1_lambda_[0.01, 0.09]	erm
Australia	4.21	4.67	-5.30	-105.84	-2884.11	-18110.55	-18395.28
State	8.20	4.46	4.52	-40.55	-2018.79	-5729.73	-5772.14
Gender	13.42	8.12	10.87	-30.36	-1893.14	-5309.36	-5326.10
Legal	11.68	9.00	9.88	-6.36	-1538.18	-2578.87	-2585.93
Indigenous	12.27	9.78	9.84	0.00	-1412.38	-2149.98	-2154.93
Overall	9.09	6.78	4.34	-46.57	-2084.76	-8362.88	-8463.75

Labour

ARIMA

	mintshrink	ols	wls	bottomup	mintsample	erm	case1_lambda_[0.1, 0.9]
Total Employ	7.46	3.91	2.88	-14.45	-103.28	-9904.66	-75173.41
Main Occupa	1.54	1.61	1.03	-3.47	-115.96	-15244.29	-10316.04
Employment	4.90	3.20	3.71	2.50	-99.16	-12379.33	-5778.02
Gender	1.07	0.42	0.03	0.00	-116.41	-11541.61	-3781.49
Overall	3.45	2.18	1.90	-2.76	-108.89	-12542.81	-18673.49

ETS

	mintshrink	ols	wls	bottomup	mintsample	erm	case1_lambd
Total Employ	7.46	3.91	2.88	-14.45	-103.28	-9904.66	-75173.41
Main Occupa	1.54	1.61	1.03	-3.47	-115.96	-15244.29	-10316.04
Employment	4.90	3.20	3.71	2.50	-99.16	-12379.33	-5778.02
Gender	1.07	0.42	0.03	0.00	-116.41	-11541.61	-3781.49
Overall	3.45	2.18	1.90	-2.76	-108.89	-12542.81	-18673.49

Tourism

ARIMA

	ols	mintsample	mintshrink	wls	bottomup	case1_lambda_[1, 4]
Australia	-2.49	-9.70	-30.11	-54.42	-122.15	-183.50
States	9.11	10.65	5.12	-3.14	-31.01	-46.77
Regions	5.10	4.16	7.49	6.36	0.00	-9.36
Overall	2.50	-0.86	-12.34	-27.47	-71.71	-109.47

ETS

	ols	mintsample	mintshrink	wls	bottomup	case1_lambda_[1, 4]
Australia	-1.23	-1.35	-23.67	-29.27	-59.15	-131.81
States	3.37	-0.14	-4.28	-6.43	-20.40	-49.64
Regions	3.33	-1.26	3.73	3.37	0.00	-15.98
Overall	0.80	-0.99	-13.78	-17.56	-38.68	-89.94

Wikipedia

ARIMA

	mintsample	wls	bottomup	ols	case1_lambda_1	case1_lambda_[0.1, 0.9]	case1_lambda_[1, 4]	case1_lambda_[0.01, 0.09]	erm	mintshrink
Total	14.96	11.85	7.61	4.37	12.10	2.97	-26.27	-100.48	-149.06	-923.18
Access	18.98	14.15	8.59	6.29	12.31	5.92	-8.79	-42.47	-78.57	-600.83
Agent	8.90	3.62	-2.67	-2.54	3.64	-8.43	-19.93	-58.17	-98.46	-620.11
Language	14.41	8.75	3.90	1.55	4.62	-19.55	-20.31	-66.59	-90.72	-643.66
Purpose	11.02	7.05	6.26	0.14	-42.55	-31.38	-36.37	-53.84	-58.37	-921.73
Article	4.35	1.34	0.00	0.02	-69.99	-70.65	-70.54	-72.28	-72.35	-849.05
Overall	13.10	8.71	4.39	2.08	-4.97	-13.16	-25.67	-66.18	-96.71	-744.63

ETS

	mintsample	wls	bottomup	ols	erm	case1_lambda_[0.01, 0.09]	case1_lambda_[1, 4]	case1_lambda_1	case1_lambda_[0.1, 0.9]	mintshrink
Total	0.78	0.37	-0.42	-1.87	-81.33	-83.42	-91.70	-94.52	-124.29	-4629.76
Access	5.46	5.19	4.94	2.82	-42.12	-42.03	-47.45	-48.19	-58.34	-3481.03
Agent	5.78	5.49	4.55	2.34	-42.58	-42.21	-46.36	-49.02	-57.79	-3518.68
Language	1.90	1.55	0.75	-0.22	-57.90	-56.84	-60.11	-64.18	-74.22	-3428.47
Purpose	-1.70	-1.93	-3.31	-1.07	-52.80	-52.12	-54.02	-55.34	-60.16	-4194.51
Article	1.80	1.30	0.00	0.54	-67.22	-67.23	-67.28	-67.37	-67.68	-4161.60
Overall	2.74	2.41	1.57	0.56	-56.79	-56.90	-61.34	-63.43	-75.75	-3881.10

Running Time for ML method

This table shows the average running time per rolling window when training and executing a trained ML reconciliation method. The average training time also includes the hyper-parameter tuning time.

Dataset	**Base Model**	**Average training time**	**Average execution time once trained**
Prison	ARIMA	25.47 minutes	0.024 minutes
	ETS	22.22 minutes	0.024 minutes
Labour	ARIMA	15.65 minutes	0.023 minutes
	ETS	15.09 minutes	0.023 minutes
Tourism	ARIMA	11.46 minutes	0.028 minutes

	ETS	11.79 minutes	0.023 minutes
Wikipedia	ARIMA	77.79 minutes	0.033 minutes
	ETS	72.34 minutes	0.035 minutes