Supplementary Appendix: A Novel Approach to Predictive Accuracy Testing in Nested Environments

Jean-Yves Pitarakis*

University of Southampton

Department of Economics

j.pitarakis@soton.ac.uk

April 17, 2023

1. Introduction

This supplementary appendix provides further simulation results illustrating the size and power properties of the two test statistics $\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$ and $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ and their adjusted versions $\mathcal{S}_{T,adj}^0(\lambda_1^0, \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0; \lambda_2^0)$ introduced in the paper. Emphasis is placed on robustness considerations as well as additional DGPs designed to capture commonly encountered stylised facts in applied work (e.g., presence of unit-roots and cointegration amongst predictors).

The structure of this supplementary appendix is as follows:

Section 2 considers the same DGPs and simulations as in the main text and repeats all size and power experiments allowing for conditional heteroskedasticity and the use of Newey-West

*Address for Correspondence: Jean-Yves Pitarakis, Department of Economics, University of Southampton, Southampton SO17 1BJ, United-Kingdom.

type adjustments in the variance normalisers of the test statistics. We also use this set of additional experiments to document the impact of using alternative magnitudes for τ_0 in the implementation of the average based statistics $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0; \lambda_2^0)$.

Section 3 revisits the local power experiments of the main text (i.e., Tables 3-4) considering smaller departures from the null hypothesis. This is motivated by the fact that under highly persistent predictors (e.g., $\phi_1 = 0.95, 0.98$) our empirical power estimates quickly reached magnitudes such as 99%-100% making it important to also gauge power for magnitudes that are more local to zero.

Section 4 focuses on the sensitivity of outcomes to the sample size via an alternative choice for $k_0 = [T\pi_0]$. We may recall that k_0 is the starting point of the recursive estimation procedure so that the effective size for the sample of forecast errors is $T - k_0$. All our experiments in the main text were implemented using $\pi_0 = 0.25$. Here we repeat the same experiments using $\pi_0 = 0.5$ instead so that the effective sample size under $T = 250$ for instance is now $(T - k_0) = 125$.

Finally, Section 5 introduces an additional DGP designed to capture additional features commonly encountered in applied work (e.g., exact unit-roots and cointegration among predictors).

2. Size and Power under Conditional Heteroskedasticity

Our analysis continues to be based on the two DGPs (DGP1 and DGP2) whose features are as described in the main text. The emphasis here is on documenting performance under conditional heteroskedasticity when using Newey-West type adjustments for the variance normalisers of the statistics. Recall from (8) and (9) and Assumption A that $\hat{\sigma}^2$ refers to a consistent estimator of the variance of $\eta_t = u_{t+1}^2 - \sigma_u^2$. Our analysis under conditional homoskedasticity used $\hat{\sigma}_{hom}^2 = \sum_t \hat{\eta}_{t+1}^2/(T - k_0)$ while in what follows we use a Newey-West type formulation as $\hat{\sigma}_{nw}^2 = \hat{\gamma}_0 + 2\sum_\ell k(\ell/M_T)\hat{\gamma}_\ell$ where $\hat{\gamma}_\ell = \sum_t \hat{\eta}_{t+1}\hat{\eta}_{t+1-\ell}/(T - k_0)$. Here $k(.)$ denotes a kernel function (set at Parzen's kernel) and M the bandwidth satisfying $M \to \infty$ and $M/T \to 0$. Besides addressing the issue of robustness to conditional heteroskedasticity our additional simulations also consider an extended set of parameterisations for the user-inputs required when implementing our proposed inferences. All results are based on a 10% nominal size and rely on 10000 replications.

Size Properties: DGP1 and DGP2 under conditional heteroskedasticity

Table S1 presents empirical size outcomes for $\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0, \lambda_2^0)$. We note that virtually all outcomes fall in the vicinity of the 10% nominal size suggesting that the NW type adjustment applied to these test statistics is able to accurately neutralise the underlying heteroskedasticity (albeit with some mild overrejections in small samples for $\mathcal{S}_{T,adj}^0(\lambda_1^0, \lambda_2^0)$ which are quickly resorbed as T exceeds 500).

Tables S2 and S3 repeat the same exercise for the average based statistics $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0; \lambda_2^0)$ across $\tau_0 \in \{0.5, 0.8\}$. Outcomes continue to indicate good size control across alternative parameterisations. The adjusted versions of the statistics show a tendency to overreject under $T = 250$ and ϕ close to 1. These overrejections are again quickly resorbed as T is allowed to grow. Finally, we note little difference in size outcomes between the implementation of $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ or $\overline{\mathcal{S}}_{T,adj}(\tau_0; \lambda_2^0)$ using either $\tau_0 = 0.5$ and $\tau_0 = 0.8$.

At this stage it is also interesting to contrast the outcomes in Tables S2 and S3 which differ in the magnitude of τ_0 used in the implementation of the average based statistics. We may recall that τ_0 determines the range of λ_1's used in the averaging of the $Z_T(\lambda_1, \lambda_2^0)$ statistic (see equation (9) in the main text) with larger magnitudes of τ_0 indicating an averaging across the larger magnitudes of λ_1. We may also recall that our local power analysis implied more favorable power outcomes for τ_0 near one combined with $\lambda_2^0 = 0.5\tau_0 + 0.5$. Comparing the size estimates of $\overline{\mathcal{S}}_{T,adj}(\tau_0; \lambda_2^0)$ in Table S2 based on $\tau_0 = 0.5$ with Table S3 based on $\tau_0 = 0.8$ we note greater size distortions that occur for the $\tau_0 = 0.8$ implementation, albeit solely for $T = 250$ and $\lambda_2^0 = 0.85$ and/or $\lambda_2^0 = 0.90$. This is clearly not surprising given our earlier point about a power maximising configuration of $(\tau_0; \lambda_2^0 = 0.5\tau_0 + 0.5)$. Nevertheless we can also note that the documented oversizeness nearly disappears under $T = 1000$ in addition to being mainly confined to DGPs with highly persistent predictors (e.g., $\phi = 0.98$).

Next, in Tables S4-S6 we repeat the size experiments associated with DGP2 under conditional

Table S1: **DGP1** Empirical size of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ under conditional heteroskedasticity

	λ_2^0	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	
						$\phi = 0.75$						DM_{nw}
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$	T=250	0.101	0.106	0.111	0.110	0.112	0.114	0.109	0.107	0.102	0.080	0.008
	T=500	0.102	0.105	0.113	0.117	0.122	0.120	0.118	0.119	0.108	0.091	0.007
	T=1000	0.108	0.111	0.113	0.113	0.114	0.116	0.118	0.116	0.114	0.097	0.007
												CW_{nw}
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$	T=250	0.120	0.123	0.130	0.129	0.133	0.135	0.129	0.132	0.134	0.118	0.060
	T=500	0.112	0.119	0.124	0.132	0.134	0.135	0.135	0.133	0.129	0.116	0.059
	T=1000	0.115	0.117	0.121	0.120	0.123	0.126	0.128	0.126	0.127	0.113	0.058
						$\phi = 0.95$						DM_{nw}
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$	T=250	0.090	0.097	0.098	0.101	0.105	0.104	0.102	0.099	0.093	0.075	0.008
	T=500	0.099	0.103	0.107	0.110	0.116	0.113	0.110	0.106	0.099	0.086	0.008
	T=1000	0.097	0.101	0.106	0.107	0.115	0.114	0.115	0.112	0.107	0.096	0.007
												CW_{nw}
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$	T=250	0.111	0.121	0.123	0.126	0.134	0.135	0.133	0.131	0.129	0.120	0.071
	T=500	0.113	0.116	0.121	0.125	0.130	0.130	0.127	0.125	0.119	0.111	0.064
	T=1000	0.103	0.108	0.112	0.113	0.123	0.122	0.124	0.123	0.120	0.110	0.059
						$\phi = 0.98$						DM_{nw}
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$	T=250	0.095	0.101	0.104	0.106	0.108	0.105	0.107	0.105	0.097	0.079	0.011
	T=500	0.095	0.100	0.106	0.106	0.108	0.109	0.113	0.107	0.099	0.080	0.009
	T=1000	0.103	0.104	0.104	0.109	0.107	0.110	0.112	0.109	0.107	0.091	0.008
												CW_{nw}
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$	T=250	0.126	0.132	0.137	0.140	0.142	0.144	0.146	0.145	0.142	0.142	0.091
	T=500	0.112	0.119	0.122	0.125	0.128	0.128	0.133	0.133	0.123	0.112	0.076
	T=1000	0.111	0.113	0.114	0.120	0.117	0.119	0.125	0.121	0.120	0.109	0.060

Table S2: **DGP1** Empirical size of $\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)$ under conditional heteroskedasticity

	λ_2^0	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	1.000	
							$\phi = 0.75$						DM_{nw}
$\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)_{nw}$	T=250	0.102	0.104	0.103	0.094	0.079	0.068	0.049	0.052	0.057	0.065	0.070	0.008
	T=500	0.105	0.104	0.112	0.105	0.090	0.074	0.065	0.069	0.075	0.076	0.083	0.007
	T=1000	0.108	0.113	0.112	0.102	0.091	0.078	0.069	0.069	0.074	0.079	0.083	0.007
													CW_{nw}
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)_{nw}$	T=250	0.126	0.135	0.138	0.133	0.127	0.119	0.106	0.102	0.101	0.103	0.105	0.060
	T=500	0.120	0.124	0.135	0.131	0.120	0.106	0.099	0.101	0.104	0.098	0.102	0.059
	T=1000	0.117	0.124	0.125	0.120	0.110	0.101	0.091	0.088	0.091	0.093	0.095	0.058
							$\phi = 0.95$						DM_{nw}
$\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)_{nw}$	T=250	0.094	0.095	0.095	0.088	0.074	0.057	0.044	0.045	0.047	0.058	0.063	0.008
	T=500	0.101	0.099	0.100	0.096	0.091	0.081	0.062	0.060	0.066	0.072	0.076	0.008
	T=1000	0.099	0.102	0.100	0.100	0.094	0.078	0.075	0.077	0.084	0.084	0.092	0.007
													CW_{nw}
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)_{nw}$	T=250	0.124	0.131	0.140	0.143	0.135	0.120	0.106	0.105	0.097	0.106	0.104	0.071
	T=500	0.120	0.120	0.125	0.125	0.125	0.120	0.097	0.099	0.098	0.101	0.100	0.064
	T=1000	0.109	0.112	0.112	0.117	0.110	0.097	0.097	0.097	0.102	0.099	0.104	0.059
							$\phi = 0.98$						DM_{nw}
$\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)_{nw}$	T=250	0.095	0.100	0.097	0.085	0.076	0.057	0.047	0.041	0.049	0.056	0.057	0.011
	T=500	0.105	0.098	0.099	0.093	0.081	0.064	0.058	0.058	0.067	0.071	0.080	0.009
	T=1000	0.099	0.102	0.103	0.099	0.088	0.078	0.070	0.072	0.080	0.087	0.088	0.008
													CW_{nw}
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)_{nw}$	T=250	0.137	0.146	0.153	0.149	0.151	0.141	0.125	0.119	0.112	0.113	0.108	0.091
	T=500	0.126	0.127	0.132	0.128	0.122	0.110	0.105	0.102	0.102	0.100	0.106	0.076
	T=1000	0.110	0.115	0.120	0.118	0.110	0.104	0.096	0.096	0.101	0.102	0.102	0.060

Table S3: **DGP1** Empirical size of $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$ under conditional heteroskedasticity

	λ_2^0	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	1.000	
							$\phi = 0.75$						DM_{nw}
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)_{nw}$	T=250	0.103	0.113	0.113	0.111	0.109	0.109	0.096	0.091	0.066	0.028	0.030	0.008
	T=500	0.105	0.110	0.118	0.116	0.113	0.110	0.106	0.097	0.070	0.036	0.044	0.007
	T=1000	0.110	0.112	0.114	0.112	0.108	0.110	0.101	0.095	0.077	0.049	0.058	0.007
													CW_{nw}
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)_{nw}$	T=250	0.122	0.132	0.135	0.134	0.135	0.136	0.131	0.143	0.141	0.097	0.084	0.060
	T=500	0.116	0.121	0.133	0.131	0.126	0.128	0.128	0.130	0.126	0.084	0.082	0.059
	T=1000	0.116	0.120	0.124	0.122	0.119	0.120	0.116	0.117	0.107	0.081	0.079	0.058
							$\phi = 0.95$						DM_{nw}
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)_{nw}$	T=250	0.096	0.100	0.103	0.104	0.105	0.099	0.089	0.083	0.057	0.023	0.030	0.008
	T=500	0.104	0.108	0.111	0.109	0.114	0.115	0.103	0.089	0.059	0.030	0.038	0.008
	T=1000	0.100	0.105	0.105	0.110	0.113	0.102	0.104	0.097	0.071	0.045	0.056	0.007
													CW_{nw}
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)_{nw}$	T=250	0.118	0.126	0.134	0.135	0.139	0.135	0.135	0.148	0.147	0.104	0.090	0.071
	T=500	0.118	0.119	0.125	0.124	0.132	0.133	0.129	0.128	0.114	0.081	0.074	0.064
	T=1000	0.107	0.112	0.113	0.121	0.121	0.113	0.117	0.117	0.107	0.077	0.077	0.059
							$\phi = 0.98$						DM_{nw}
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)_{nw}$	T=250	0.100	0.104	0.111	0.104	0.104	0.101	0.093	0.086	0.060	0.026	0.025	0.011
	T=500	0.101	0.099	0.105	0.102	0.103	0.100	0.101	0.090	0.062	0.026	0.040	0.009
	T=1000	0.103	0.106	0.106	0.109	0.107	0.102	0.103	0.096	0.069	0.044	0.056	0.008
													CW_{nw}
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)_{nw}$	T=250	0.132	0.137	0.145	0.139	0.146	0.147	0.154	0.168	0.175	0.129	0.103	0.091
	T=500	0.119	0.117	0.127	0.124	0.125	0.123	0.134	0.140	0.133	0.090	0.085	0.076
	T=1000	0.111	0.115	0.115	0.119	0.118	0.116	0.120	0.122	0.104	0.079	0.080	0.060

heteroskedasticity. Outcomes continue to corroborate the fact that all of these test statistics show good to excellent size control regardless of the chosen user-inputs for $(\lambda_1^0, \lambda_2^0)$ and (τ_0, λ_2^0). An exception to this occurs for the unadjusted statistics $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$ under smaller sample sizes and λ_2^0 in the vicinity of 0.8-0.9 as it was the case in our homoskedastic experiments. Nevertheless, these distortions are nearly eliminated as we increase the sample size.

Table S4: **DGP2** Empirical size of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ under conditional heteroskedasticity

	λ_2^0	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	
												DM_{nw}
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$	T=250	0.075	0.079	0.080	0.083	0.084	0.083	0.079	0.080	0.078	0.057	0.003
	T=500	0.084	0.087	0.091	0.094	0.098	0.095	0.096	0.092	0.086	0.068	0.002
	T=1000	0.095	0.097	0.098	0.099	0.104	0.102	0.107	0.100	0.097	0.082	0.002
												CW_{nw}
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$	T=250	0.112	0.120	0.128	0.133	0.135	0.136	0.134	0.137	0.143	0.140	0.075
	T=500	0.108	0.114	0.120	0.123	0.129	0.128	0.128	0.129	0.122	0.113	0.064
	T=1000	0.110	0.112	0.115	0.118	0.122	0.124	0.128	0.124	0.125	0.114	0.067

Power Properties: DGP1 and DGP2 under conditional heteroskedasticity

Table S5: **DGP2** Empirical size of $\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)$ under conditional heteroskedasticity

	λ_2^0	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	1.000	
													DM_{nw}
$\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)_{nw}$	T=250	0.072	0.071	0.072	0.061	0.047	0.035	0.023	0.024	0.028	0.037	0.042	0.003
	T=500	0.086	0.087	0.083	0.072	0.067	0.049	0.040	0.041	0.050	0.052	0.060	0.002
	T=1000	0.092	0.090	0.088	0.080	0.072	0.059	0.052	0.057	0.061	0.069	0.073	0.002
													CW_{nw}
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)_{nw}$	T=250	0.125	0.137	0.152	0.153	0.150	0.145	0.136	0.126	0.120	0.120	0.117	0.075
	T=500	0.121	0.129	0.132	0.129	0.127	0.118	0.109	0.107	0.108	0.107	0.103	0.064
	T=1000	0.113	0.114	0.117	0.118	0.116	0.106	0.099	0.095	0.098	0.100	0.100	0.067

Table S6: **DGP2** Empirical size of $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$ under conditional heteroskedasticity

	λ_2^0	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	1.000	
													DM_{nw}
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)_{nw}$	T=250	0.076	0.083	0.083	0.083	0.080	0.075	0.063	0.055	0.036	0.015	0.014	0.003
	T=500	0.091	0.092	0.097	0.095	0.092	0.089	0.079	0.063	0.037	0.016	0.022	0.003
	T=1000	0.094	0.094	0.096	0.097	0.096	0.091	0.087	0.074	0.052	0.026	0.038	0.003
													CW_{nw}
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)_{nw}$	T=250	0.118	0.131	0.136	0.137	0.139	0.140	0.147	0.173	0.188	0.147	0.114	0.075
	T=500	0.115	0.121	0.128	0.126	0.127	0.127	0.129	0.137	0.136	0.106	0.087	0.064
	T=1000	0.111	0.113	0.117	0.118	0.119	0.116	0.119	0.123	0.119	0.088	0.080	0.067

Table S7 presents empirical power estimates for the Newey-West corrected versions of $\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0, \lambda_2^0)$ under DGP1 while Tables S8-S9 repeat the same exercise for $\overline{\mathcal{S}}_{T,adj}(\tau_0;\ \lambda_2^0)$ and $\tau_0 \in \{0.5, 0.8\}$.

In Table S7 we continue to observe the tendency of power to converge towards 1 as β moves further away from the null. Relative to the homoskedastic setting however (see Table 3 in the main text) we can also note an overall deterioration of power under $\phi_1 = 0.75$ in particular. This of course is a well known phenomenon caused by the use of the Newey-West correction which affects the rate of divergence of all test statistics under the alternative.

Tables S8 and S9 consider the power properties of the average based test statistics under $\tau_0 = 0.5$ and $\tau_0 = 0.8$ respectively. As discussed above we continue to observe a drop in empirical power for these heteroskedasticity adjusted statistics. Nevertheless all empirical powers of $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$ lie above 90% for $\phi_1 = 0.95$ and $\phi_1 = 0.98$.

Tables S10-S12 present the power properties related to DGP2 when conditional heteroskedasticity is present. Focusing first on Table S10 we note that the power outcomes are in close agreement with our homoskedasticity based experiments in the main text. Power quickly reaches

Table S7: **DGP1** Empirical Power of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ under conditional heteroskedasticity

β	-1.500	-1.750	-2.000	-2.250	-2.500	-3.000	-3.500
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$				$\phi_1 = 0.75$			
$\lambda_2^0 = 0.80$	0.175	0.200	0.235	0.263	0.308	0.414	0.545
$\lambda_2^0 = 0.85$	0.178	0.214	0.249	0.295	0.342	0.466	0.605
$\lambda_2^0 = 0.90$	0.180	0.229	0.271	0.334	0.397	0.548	0.690
$\lambda_2^0 = 0.95$	0.205	0.275	0.352	0.439	0.515	0.692	0.823
DM_{nw}	0.226	0.327	0.443	0.547	0.638	0.791	0.888
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$							
$\lambda_2^0 = 0.80$	0.267	0.323	0.397	0.474	0.545	0.703	0.831
$\lambda_2^0 = 0.85$	0.285	0.358	0.441	0.525	0.608	0.766	0.876
$\lambda_2^0 = 0.90$	0.321	0.412	0.506	0.605	0.690	0.839	0.926
$\lambda_2^0 = 0.95$	0.403	0.516	0.628	0.737	0.810	0.920	0.971
CW_{nw}	0.640	0.766	0.859	0.922	0.958	0.988	0.997
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$				$\phi_1 = 0.95$			
$\lambda_2^0 = 0.80$	0.438	0.552	0.658	0.765	0.833	0.929	0.969
$\lambda_2^0 = 0.85$	0.493	0.611	0.716	0.816	0.877	0.955	0.980
$\lambda_2^0 = 0.90$	0.565	0.694	0.787	0.875	0.922	0.972	0.991
$\lambda_2^0 = 0.95$	0.697	0.809	0.879	0.935	0.965	0.990	0.997
DM_{nw}	0.828	0.908	0.942	0.973	0.984	0.996	0.999
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$							
$\lambda_2^0 = 0.80$	0.692	0.799	0.871	0.931	0.960	0.990	0.997
$\lambda_2^0 = 0.85$	0.743	0.844	0.903	0.953	0.976	0.995	0.998
$\lambda_2^0 = 0.90$	0.808	0.897	0.943	0.972	0.987	0.997	0.999
$\lambda_2^0 = 0.95$	0.885	0.948	0.972	0.989	0.996	0.999	1.000
CW_{nw}	0.979	0.993	0.997	0.999	1.000	1.000	1.000
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$				$\phi_1 = 0.98$			
$\lambda_2^0 = 0.80$	0.709	0.815	0.883	0.931	0.960	0.987	0.995
$\lambda_2^0 = 0.85$	0.762	0.857	0.915	0.952	0.972	0.993	0.997
$\lambda_2^0 = 0.90$	0.822	0.899	0.946	0.971	0.985	0.996	0.999
$\lambda_2^0 = 0.95$	0.892	0.948	0.976	0.987	0.995	0.999	1.000
DM_{nw}	0.950	0.978	0.990	0.995	0.997	0.999	1.000
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$							
$\lambda_2^0 = 0.80$	0.883	0.939	0.971	0.985	0.993	0.998	1.000
$\lambda_2^0 = 0.85$	0.913	0.959	0.980	0.991	0.996	0.999	1.000
$\lambda_2^0 = 0.90$	0.941	0.974	0.989	0.995	0.998	1.000	1.000
$\lambda_2^0 = 0.95$	0.969	0.987	0.995	0.999	0.999	1.000	1.000
CW_{nw}	0.995	0.999	1.000	1.000	1.000	1.000	1.000

Table S8: **DGP1** Empirical Power of $\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)$ under conditional heteroskedasticity

β	-1.500	-1.750	-2.000	-2.250	-2.500	-3.000	-3.500
$\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)_{nw}$				$\phi_1 = 0.75$			
$\lambda_2^0 = 0.80$	0.212	0.267	0.338	0.417	0.487	0.627	0.759
$\lambda_2^0 = 0.85$	0.193	0.245	0.310	0.385	0.458	0.602	0.720
$\lambda_2^0 = 0.90$	0.180	0.231	0.280	0.352	0.412	0.550	0.674
$\lambda_2^0 = 0.95$	0.171	0.209	0.263	0.318	0.382	0.502	0.630
$\lambda_2^0 = 1$	0.166	0.199	0.245	0.291	0.355	0.467	0.590
DM_{nw}	0.226	0.327	0.443	0.547	0.638	0.791	0.888
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)_{nw}$							
$\lambda_2^0 = 0.80$	0.488	0.568	0.663	0.749	0.812	0.903	0.958
$\lambda_2^0 = 0.85$	0.463	0.554	0.633	0.719	0.784	0.892	0.947
$\lambda_2^0 = 0.90$	0.438	0.515	0.596	0.686	0.751	0.863	0.931
$\lambda_2^0 = 0.95$	0.405	0.484	0.561	0.647	0.714	0.837	0.914
$\lambda_2^0 = 1$	0.391	0.448	0.528	0.618	0.677	0.812	0.890
CW_{nw}	0.640	0.766	0.859	0.922	0.958	0.988	0.997
$\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)_{nw}$				$\phi_1 = 0.95$			
$\lambda_2^0 = 0.80$	0.623	0.734	0.822	0.890	0.926	0.974	0.990
$\lambda_2^0 = 0.85$	0.597	0.700	0.798	0.868	0.914	0.966	0.987
$\lambda_2^0 = 0.90$	0.554	0.663	0.758	0.833	0.891	0.955	0.982
$\lambda_2^0 = 0.95$	0.516	0.623	0.724	0.801	0.864	0.941	0.975
$\lambda_2^0 = 1$	0.479	0.587	0.687	0.771	0.838	0.925	0.968
DM_{nw}	0.828	0.908	0.942	0.973	0.984	0.996	0.999
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)_{nw}$							
$\lambda_2^0 = 0.80$	0.835	0.904	0.947	0.971	0.984	0.997	0.999
$\lambda_2^0 = 0.85$	0.811	0.885	0.936	0.965	0.981	0.996	0.998
$\lambda_2^0 = 0.90$	0.787	0.866	0.918	0.957	0.976	0.994	0.998
$\lambda_2^0 = 0.95$	0.753	0.843	0.902	0.946	0.970	0.991	0.997
$\lambda_2^0 = 1$	0.723	0.818	0.881	0.935	0.962	0.989	0.996
CW_{nw}	0.979	0.993	0.997	0.999	1.000	1.000	1.000
$\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)_{nw}$				$\phi_1 = 0.98$			
$\lambda_2^0 = 0.80$	0.833	0.903	0.944	0.968	0.981	0.995	0.998
$\lambda_2^0 = 0.85$	0.812	0.888	0.934	0.961	0.976	0.994	0.997
$\lambda_2^0 = 0.90$	0.780	0.864	0.918	0.949	0.971	0.990	0.997
$\lambda_2^0 = 0.95$	0.746	0.838	0.899	0.938	0.963	0.987	0.996
$\lambda_2^0 = 1$	0.715	0.814	0.877	0.925	0.954	0.984	0.994
DM_{nw}	0.950	0.978	0.990	0.995	0.997	0.999	1.000
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)_{nw}$							
$\lambda_2^0 = 0.80$	0.937	0.970	0.985	0.993	0.996	0.999	1.000
$\lambda_2^0 = 0.85$	0.932	0.965	0.983	0.993	0.996	0.999	1.000
$\lambda_2^0 = 0.90$	0.918	0.957	0.980	0.991	0.995	0.999	1.000
$\lambda_2^0 = 0.95$	0.904	0.950	0.976	0.989	0.994	0.998	1.000
$\lambda_2^0 = 1$	0.892	0.942	0.969	0.986	0.993	0.998	1.000
CW_{nw}	0.995	0.999	1.000	1.000	1.000	1.000	1.000

Table S9: **DGP1** Empirical Power of $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$ under conditional heteroskedasticity

β	-1.500	-1.750	-2.000	-2.250	-2.500	-3.000	-3.500
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)_{nw}$				$\phi_1 = 0.75$			
$\lambda_2^0 = 0.80$	0.209	0.247	0.308	0.371	0.439	0.584	0.732
$\lambda_2^0 = 0.85$	0.258	0.339	0.416	0.507	0.593	0.749	0.860
$\lambda_2^0 = 0.90$	0.315	0.413	0.510	0.615	0.702	0.834	0.915
$\lambda_2^0 = 0.95$	0.263	0.348	0.454	0.560	0.642	0.778	0.884
$\lambda_2^0 = 1$	0.211	0.279	0.355	0.444	0.522	0.674	0.796
DM_{nw}	0.226	0.327	0.443	0.547	0.638	0.791	0.888
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)_{nw}$							
$\lambda_2^0 = 0.80$	0.435	0.519	0.624	0.716	0.785	0.898	0.959
$\lambda_2^0 = 0.85$	0.546	0.649	0.747	0.830	0.885	0.961	0.986
$\lambda_2^0 = 0.90$	0.634	0.735	0.825	0.889	0.932	0.977	0.992
$\lambda_2^0 = 0.95$	0.584	0.688	0.783	0.858	0.905	0.961	0.984
$\lambda_2^0 = 1$	0.515	0.603	0.698	0.781	0.843	0.928	0.967
CW_{nw}	0.640	0.766	0.859	0.922	0.958	0.988	0.997
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)_{nw}$				$\phi_1 = 0.95$			
$\lambda_2^0 = 0.80$	0.597	0.716	0.805	0.882	0.925	0.976	0.991
$\lambda_2^0 = 0.85$	0.737	0.840	0.902	0.949	0.972	0.994	0.997
$\lambda_2^0 = 0.90$	0.820	0.898	0.938	0.969	0.983	0.995	0.998
$\lambda_2^0 = 0.95$	0.770	0.860	0.913	0.952	0.973	0.991	0.997
$\lambda_2^0 = 1$	0.672	0.776	0.850	0.908	0.943	0.979	0.993
DM_{nw}	0.828	0.908	0.942	0.973	0.984	0.996	0.999
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)_{nw}$							
$\lambda_2^0 = 0.80$	0.824	0.903	0.945	0.974	0.985	0.998	0.999
$\lambda_2^0 = 0.85$	0.904	0.955	0.979	0.990	0.996	0.999	1.000
$\lambda_2^0 = 0.90$	0.937	0.972	0.987	0.994	0.998	1.000	1.000
$\lambda_2^0 = 0.95$	0.916	0.961	0.978	0.991	0.996	0.999	1.000
$\lambda_2^0 = 1$	0.865	0.928	0.958	0.982	0.989	0.998	0.999
CW_{nw}	0.979	0.993	0.997	0.999	1.000	1.000	1.000
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)_{nw}$				$\phi_1 = 0.98$			
$\lambda_2^0 = 0.80$	0.827	0.907	0.943	0.971	0.986	0.995	0.999
$\lambda_2^0 = 0.85$	0.909	0.954	0.979	0.990	0.995	0.999	1.000
$\lambda_2^0 = 0.90$	0.942	0.971	0.988	0.994	0.997	0.999	1.000
$\lambda_2^0 = 0.95$	0.920	0.958	0.978	0.991	0.995	0.999	1.000
$\lambda_2^0 = 1$	0.864	0.923	0.957	0.980	0.988	0.997	0.999
DM_{nw}	0.950	0.978	0.990	0.995	0.997	0.999	1.000
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)_{nw}$							
$\lambda_2^0 = 0.80$	0.938	0.972	0.989	0.996	0.997	1.000	1.000
$\lambda_2^0 = 0.85$	0.971	0.988	0.996	0.999	0.999	1.000	1.000
$\lambda_2^0 = 0.90$	0.985	0.993	0.997	1.000	1.000	1.000	1.000
$\lambda_2^0 = 0.95$	0.977	0.990	0.995	0.998	0.999	1.000	1.000
$\lambda_2^0 = 1$	0.955	0.981	0.991	0.997	0.999	1.000	1.000
CW_{nw}	0.995	0.999	1.000	1.000	1.000	1.000	1.000

Table S10: **DGP2** Empirical power of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ under conditional heteroskedasticity

	λ_2^0	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	
												DM_{nw}
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$	T=250	0.521	0.571	0.620	0.669	0.722	0.776	0.833	0.883	0.930	0.971	0.889
	T=500	0.783	0.835	0.877	0.913	0.944	0.967	0.985	0.994	0.998	1.000	0.998
	T=1000	0.959	0.976	0.987	0.994	0.998	0.999	1.000	1.000	1.000	1.000	1.000
												CW_{nw}
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)_{nw}$	T=250	0.910	0.936	0.955	0.969	0.979	0.987	0.993	0.997	0.999	1.000	1.000
	T=500	0.992	0.996	0.998	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	T=1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table S11: **DGP2** Empirical power of $\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)$ under conditional heteroskedasticity

	λ_2^0	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	1.000	
													DM_{nw}
$\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)_{nw}$	T=250	0.528	0.612	0.692	0.768	0.818	0.840	0.837	0.816	0.785	0.751	0.714	0.854
	T=500	0.717	0.807	0.878	0.927	0.948	0.956	0.951	0.937	0.918	0.893	0.861	0.988
	T=1000	0.879	0.932	0.965	0.982	0.990	0.990	0.990	0.985	0.980	0.973	0.958	1.000
													CW_{nw}
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)_{nw}$	T=250	0.864	0.916	0.946	0.970	0.977	0.980	0.980	0.976	0.967	0.959	0.946	0.997
	T=500	0.954	0.977	0.988	0.993	0.996	0.997	0.996	0.994	0.993	0.989	0.985	1.000
	T=1000	0.989	0.994	0.998	0.999	0.999	1.000	0.999	0.999	0.999	0.998	0.997	1.000

magnitudes near 1 for inferences based on $\mathcal{S}^0_{T,adj}(\lambda_1^0 = 1, \lambda_2^0)_{nw}$. For the average based statistics in Tables S11-S12 the empirical results also clearly corroborate our theoretical analysis that allowed us to establish that power peaks for $\lambda_2^0 = 0.5\tau_0 + 0.5$ (i.e., $\lambda_2^0 = 0.75$ under $\tau_0 = 0.5$ and $\lambda_2^0 = 0.95$ under $\tau_0 = 0.8$).

Table S12: **DGP2** Empirical power of $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$ under conditional heteroskedasticity

	λ_2^0	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	1.000	
													DM_{nw}
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)_{nw}$	T=250	0.429	0.472	0.517	0.565	0.627	0.711	0.820	0.912	0.945	0.929	0.881	0.854
	T=500	0.589	0.639	0.696	0.754	0.820	0.895	0.951	0.986	0.992	0.984	0.963	0.854
	T=1000	0.759	0.810	0.858	0.903	0.944	0.973	0.992	0.998	0.999	0.998	0.992	1.000
													CW_{nw}
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)_{nw}$	T=250	0.778	0.822	0.858	0.897	0.928	0.957	0.980	0.994	0.997	0.994	0.986	0.997
	T=500	0.899	0.931	0.951	0.969	0.981	0.991	0.997	0.999	1.000	0.999	0.997	1.000
	T=1000	0.968	0.980	0.988	0.993	0.997	0.999	0.999	1.000	1.000	1.000	0.999	1.000

Recalling that these power experiments were implemented using a fixed $T = 500$ it is also useful to point out that for such sample sizes the size properties of the tests were also not far off their nominal counterparts. Overall, these power outcomes provide an encouraging assessment of the finite sample behavior of the proposed test statistics.

3. Power under additional parameterisations

Here we reconsider the power experiments associated with DGP1 in the main text (Tables 3-4) by considering parameterisations that bring the DGP closer to the null (i.e., smaller magnitudes of β). The experiments in Tables 3-4 of the main text were implemented for

$\beta \in \{-1.50, -1.75, -2.00, -2.25, -2.50, -3.00, -3.50\}$. Here we repeat the same experiments using $\beta \in \{-0.50, -0.75, -1.00, -1.25\}$ with outcomes presented in Tables S13-S15. The sample size for these local power experiments is kept at $T = 500$ as in the main text.

Table S13: **DGP1** Empirical Power of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ under conditional homoskedasticity

β	-0.50	-0.75	-1.00	-1.25
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$		$\phi_1 = 0.75$		
$\lambda_2^0 = 0.80$	0.098	0.109	0.128	0.155
$\lambda_2^0 = 0.85$	0.100	0.112	0.136	0.173
$\lambda_2^0 = 0.90$	0.104	0.113	0.146	0.188
$\lambda_2^0 = 0.95$	0.099	0.132	0.179	0.241
DM	0.025	0.062	0.118	0.204
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)_{adj}$				
$\lambda_2^0 = 0.80$	0.128	0.159	0.202	0.268
$\lambda_2^0 = 0.85$	0.136	0.165	0.229	0.304
$\lambda_2^0 = 0.90$	0.147	0.188	0.261	0.360
$\lambda_2^0 = 0.95$	0.167	0.243	0.336	0.453
CW	0.147	0.273	0.440	0.597
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$		$\phi_1 = 0.95$		
$\lambda_2^0 = 0.80$	0.124	0.194	0.323	0.464
$\lambda_2^0 = 0.85$	0.126	0.218	0.363	0.523
$\lambda_2^0 = 0.90$	0.138	0.255	0.419	0.594
$\lambda_2^0 = 0.95$	0.170	0.317	0.526	0.711
DM	0.096	0.289	0.531	0.719
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)_{adj}$				
$\lambda_2^0 = 0.80$	0.197	0.344	0.543	0.713
$\lambda_2^0 = 0.85$	0.217	0.388	0.597	0.763
$\lambda_2^0 = 0.90$	0.251	0.446	0.666	0.820
$\lambda_2^0 = 0.95$	0.321	0.542	0.755	0.886
CW	0.379	0.655	0.860	0.943
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$		$\phi_1 = 0.98$		
$\lambda_2^0 = 0.80$	0.179	0.353	0.558	0.735
$\lambda_2^0 = 0.85$	0.201	0.398	0.609	0.782
$\lambda_2^0 = 0.90$	0.231	0.458	0.674	0.833
$\lambda_2^0 = 0.95$	0.288	0.557	0.763	0.896
DM	0.229	0.551	0.773	0.905
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)_{adj}$				
$\lambda_2^0 = 0.80$	0.322	0.565	0.769	0.892
$\lambda_2^0 = 0.85$	0.358	0.618	0.805	0.918
$\lambda_2^0 = 0.90$	0.408	0.679	0.849	0.942
$\lambda_2^0 = 0.95$	0.483	0.758	0.901	0.969
CW	0.562	0.830	0.949	0.986

It is important to point out that Tables S13-S15 consider signal to noise ratios that are particularly low when compared with Tables 3-4 in the main text. Nevertheless we can still observe empirical powers that reach magnitudes as high as 80-99% for a selection of outcomes. For $\beta = -0.75$ for instance and focusing on $\phi_1 = 0.98$ we note empirical powers in the vicinity of 80% for $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0 = 0.90)$. The same frequency declines towards the 60-65% range

Table S14: **DGP1** Empirical Power of $\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)$ under conditional homoskedasticity

β	-0.50	-0.75	-1.00	-1.25
$\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)$		$\phi_1 = 0.75$		
$\lambda_2^0 = 0.80$	0.087	0.118	0.155	0.215
$\lambda_2^0 = 0.85$	0.087	0.113	0.149	0.204
$\lambda_2^0 = 0.90$	0.087	0.113	0.144	0.191
$\lambda_2^0 = 0.95$	0.088	0.113	0.134	0.175
$\lambda_2^0 = 1$	0.091	0.113	0.127	0.169
DM	0.025	0.062	0.118	0.204
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)$				
$\lambda_2^0 = 0.80$	0.150	0.210	0.294	0.393
$\lambda_2^0 = 0.85$	0.142	0.196	0.273	0.378
$\lambda_2^0 = 0.90$	0.139	0.184	0.249	0.344
$\lambda_2^0 = 0.95$	0.125	0.176	0.227	0.315
$\lambda_2^0 = 1$	0.126	0.164	0.213	0.296
CW	0.147	0.273	0.440	0.597
$\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)$		$\phi_1 = 0.95$		
$\lambda_2^0 = 0.80$	0.148	0.271	0.451	0.617
$\lambda_2^0 = 0.85$	0.141	0.261	0.420	0.584
$\lambda_2^0 = 0.90$	0.131	0.244	0.384	0.548
$\lambda_2^0 = 0.95$	0.135	0.230	0.354	0.505
$\lambda_2^0 = 1$	0.131	0.221	0.332	0.473
DM	0.096	0.289	0.531	0.719
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)$				
$\lambda_2^0 = 0.80$	0.282	0.468	0.677	0.822
$\lambda_2^0 = 0.85$	0.261	0.447	0.645	0.805
$\lambda_2^0 = 0.90$	0.246	0.421	0.616	0.775
$\lambda_2^0 = 0.95$	0.228	0.398	0.580	0.743
$\lambda_2^0 = 1$	0.215	0.373	0.551	0.719
CW	0.379	0.655	0.860	0.943
$\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0)$		$\phi_1 = 0.98$		
$\lambda_2^0 = 0.80$	0.235	0.466	0.672	0.832
$\lambda_2^0 = 0.85$	0.232	0.448	0.644	0.811
$\lambda_2^0 = 0.90$	0.212	0.421	0.618	0.782
$\lambda_2^0 = 0.95$	0.198	0.390	0.584	0.757
$\lambda_2^0 = 1$	0.189	0.368	0.551	0.727
DM	0.229	0.551	0.773	0.905
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.5; \lambda_2^0)$				
$\lambda_2^0 = 0.80$	0.419	0.672	0.839	0.937
$\lambda_2^0 = 0.85$	0.403	0.657	0.821	0.929
$\lambda_2^0 = 0.90$	0.376	0.632	0.805	0.919
$\lambda_2^0 = 0.95$	0.355	0.606	0.786	0.907
$\lambda_2^0 = 1$	0.338	0.583	0.768	0.897
CW	0.562	0.830	0.949	0.986

Table S15: **DGP1** Empirical Power of $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$ under conditional homoskedasticity

β	-0.50	-0.75	-1.00	-1.25
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$		$\phi_1 = 0.75$		
$\lambda_2^0 = 0.80$	0.101	0.122	0.155	0.203
$\lambda_2^0 = 0.85$	0.102	0.137	0.194	0.274
$\lambda_2^0 = 0.90$	0.104	0.147	0.230	0.336
$\lambda_2^0 = 0.95$	0.084	0.132	0.192	0.282
$\lambda_2^0 = 1$	0.086	0.117	0.159	0.233
DM	0.025	0.062	0.118	0.204
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$				
$\lambda_2^0 = 0.80$	0.152	0.206	0.277	0.375
$\lambda_2^0 = 0.85$	0.185	0.257	0.370	0.499
$\lambda_2^0 = 0.90$	0.215	0.303	0.437	0.579
$\lambda_2^0 = 0.95$	0.176	0.265	0.389	0.524
$\lambda_2^0 = 1$	0.151	0.218	0.312	0.431
CW	0.147	0.273	0.440	0.597
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$		$\phi_1 = 0.95$		
$\lambda_2^0 = 0.80$	0.144	0.266	0.445	0.608
$\lambda_2^0 = 0.85$	0.177	0.350	0.565	0.739
$\lambda_2^0 = 0.90$	0.209	0.414	0.637	0.799
$\lambda_2^0 = 0.95$	0.182	0.375	0.575	0.750
$\lambda_2^0 = 1$	0.157	0.297	0.486	0.659
DM	0.096	0.289	0.531	0.719
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$				
$\lambda_2^0 = 0.80$	0.270	0.459	0.680	0.821
$\lambda_2^0 = 0.85$	0.336	0.574	0.778	0.895
$\lambda_2^0 = 0.90$	0.402	0.634	0.828	0.927
$\lambda_2^0 = 0.95$	0.354	0.593	0.788	0.903
$\lambda_2^0 = 1$	0.297	0.511	0.712	0.851
CW	0.379	0.655	0.860	0.943
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$		$\phi_1 = 0.98$		
$\lambda_2^0 = 0.80$	0.240	0.462	0.685	0.840
$\lambda_2^0 = 0.85$	0.310	0.582	0.783	0.907
$\lambda_2^0 = 0.90$	0.360	0.651	0.830	0.934
$\lambda_2^0 = 0.95$	0.317	0.603	0.793	0.913
$\lambda_2^0 = 1$	0.265	0.519	0.720	0.865
DM	0.229	0.551	0.773	0.905
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$				
$\lambda_2^0 = 0.80$	0.421	0.676	0.850	0.943
$\lambda_2^0 = 0.85$	0.505	0.768	0.909	0.970
$\lambda_2^0 = 0.90$	0.567	0.811	0.927	0.981
$\lambda_2^0 = 0.95$	0.521	0.783	0.913	0.975
$\lambda_2^0 = 1$	0.457	0.718	0.872	0.957
CW	0.562	0.830	0.949	0.986

for less persistent predictors. Finally, it is also noteworthy to point out that inferences based on $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0 = 0.9)$ (as suggested by our local power analysis) lead to empirical powers similar to those displayed by the CW statistic, despite the latter being severely undersized. If we focus on Table S15 under $\phi_1 = 0.75$ for instance, inferences based on $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0 = 0.9)$ led to power estimates of $(21.5\%, 30.3\%, 43.7\%, 57.9\%)$ compared with $(14.7\%, 27.3\%, 44.0\%, 59.7\%)$ for CW.

4. Size and Power under $\pi_0 = 0.50$

Here we repeat our size and power based experiments presented in the main text using an alternative starting point that initiates the recursive forecasts. Recalling that estimation starts at $k_0 = [T\pi_0]$ so that the effective sample size is given by $T - k_0$, our simulations in the main text were implemented using $\pi_0 = 0.25$ throughout. In the Tables that follow we consider the case of $\pi_0 = 0.50$ instead. This is also meant to document the behavior of our test statistics under smaller samples sizes. Under $T = 250$ for instance, the effective number of forecast errors is about 125.

Tables S16-S17 focus on size under DGP1 and can be compared with Tables 1-2 in the main text while Tables S18-S19 correspond to DGP2 and can be compared with Tables 5-6 in the main text. On the basis of these comparisons we can clearly argue that the use of $\pi_0 = 0.50$ instead of $\pi_0 = 0.25$ does not cause any significant differences in the empirical size estimates of both test statistics and across all sample size configurations. Focusing on $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0 = 1)$ under $\phi_1 = 0.75$ for instance we had a size estimate of 11.0% under $(T, \pi_0) = (250, 0.25)$ compared with 10.2% under $(T, \pi_0) = (250, 0.50)$ in Table S17. These results indicate that finite sample approximations typically remain accurate across most parameterisations.

Naturally, the picture is mildly different when it comes to power. Contrasting Table S20 with Table 3 in the main text we note a drop in power of about ten percentage points overall. Given a drop of 25% in the effective sample size this comes across as a very reasonably sized loss. Similar orders of magnitude also characterise $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$ when comparing Table S21 with Table 4 in the main text. For DGP2 the spreads in power are even smaller when contrasting Table

Table S16: **DGP1** Empirical size of $\mathcal{S}^0_T(\lambda^0_1 = 1, \lambda^0_2)$ and $\mathcal{S}^0_{T,adj}(\lambda^0_1 = 1, \lambda^0_2)$ under conditional homoskedasticity and 10% nominal size ($\pi_0 = 0.50$)

	λ^0_2	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	
							$\phi = 0.75$					DM
$\mathcal{S}^0_T(\lambda^0_1 = 1, \lambda^0_2)$	T=250	0.095	0.096	0.097	0.096	0.095	0.096	0.095	0.097	0.092	0.090	0.019
	T=500	0.100	0.099	0.102	0.099	0.100	0.103	0.103	0.097	0.101	0.089	0.017
	T=1000	0.094	0.095	0.094	0.093	0.094	0.097	0.101	0.097	0.095	0.092	0.018
												CW
$\mathcal{S}^0_{T,adj}(\lambda^0_1 = 1, \lambda^0_2)$	T=250	0.107	0.107	0.107	0.108	0.110	0.110	0.112	0.116	0.114	0.126	0.061
	T=500	0.106	0.106	0.110	0.107	0.111	0.112	0.113	0.112	0.116	0.114	0.060
	T=1000	0.098	0.098	0.100	0.098	0.101	0.105	0.108	0.107	0.108	0.110	0.063
							$\phi = 0.95$					DM
$\mathcal{S}^0_T(\lambda^0_1 = 1, \lambda^0_2)$	T=250	0.094	0.092	0.090	0.085	0.089	0.087	0.088	0.089	0.085	0.086	0.022
	T=500	0.094	0.099	0.098	0.100	0.094	0.094	0.092	0.091	0.092	0.085	0.021
	T=1000	0.091	0.091	0.091	0.095	0.098	0.098	0.095	0.100	0.099	0.095	0.020
												CW
$\mathcal{S}^0_{T,adj}(\lambda^0_1 = 1, \lambda^0_2)$	T=250	0.106	0.105	0.105	0.101	0.106	0.109	0.108	0.114	0.114	0.125	0.070
	T=500	0.103	0.109	0.109	0.112	0.105	0.107	0.106	0.108	0.113	0.114	0.068
	T=1000	0.095	0.097	0.097	0.102	0.105	0.106	0.105	0.111	0.110	0.111	0.065
							$\phi = 0.98$					DM
$\mathcal{S}^0_T(\lambda^0_1 = 1, \lambda^0_2)$	T=250	0.090	0.095	0.091	0.088	0.090	0.090	0.092	0.092	0.087	0.092	0.034
	T=500	0.093	0.094	0.093	0.095	0.094	0.093	0.090	0.093	0.088	0.083	0.026
	T=1000	0.094	0.097	0.096	0.097	0.099	0.095	0.096	0.100	0.092	0.092	0.022
												CW
$\mathcal{S}^0_{T,adj}(\lambda^0_1 = 1, \lambda^0_2)$	T=250	0.108	0.115	0.114	0.110	0.115	0.118	0.123	0.131	0.127	0.151	0.094
	T=500	0.103	0.106	0.107	0.109	0.109	0.110	0.110	0.115	0.115	0.120	0.078
	T=1000	0.100	0.104	0.102	0.103	0.107	0.105	0.105	0.112	0.108	0.113	0.071

Table S17: **DGP1** Empirical size of $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda^0_2)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda^0_2)$ under conditional homoskedasticity and 10% nominal size; ($\pi_0 = 0.50$)

	λ^0_2	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00	
							$\phi = 0.75$						DM
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda^0_2)$	T=250	0.093	0.098	0.097	0.097	0.097	0.097	0.097	0.104	0.084	0.058	0.063	0.019
	T=500	0.098	0.098	0.104	0.098	0.100	0.098	0.100	0.097	0.083	0.074	0.072	0.017
	T=1000	0.095	0.094	0.094	0.095	0.094	0.094	0.097	0.098	0.081	0.071	0.080	0.018
													CW
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda^0_2)$	T=250	0.104	0.108	0.110	0.109	0.116	0.117	0.127	0.150	0.143	0.114	0.102	0.061
	T=500	0.104	0.107	0.111	0.108	0.112	0.112	0.122	0.132	0.126	0.110	0.096	0.060
	T=1000	0.100	0.097	0.100	0.102	0.101	0.103	0.112	0.118	0.109	0.095	0.099	0.063
							$\phi = 0.95$						DM
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda^0_2)$	T=250	0.094	0.093	0.093	0.087	0.091	0.093	0.091	0.097	0.081	0.060	0.060	0.022
	T=500	0.097	0.098	0.101	0.104	0.100	0.092	0.097	0.089	0.081	0.063	0.068	0.021
	T=1000	0.092	0.092	0.089	0.094	0.098	0.098	0.096	0.098	0.084	0.070	0.075	0.020
													CW
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda^0_2)$	T=250	0.109	0.111	0.110	0.107	0.113	0.124	0.128	0.152	0.154	0.121	0.103	0.070
	T=500	0.106	0.108	0.113	0.115	0.115	0.110	0.118	0.124	0.127	0.106	0.098	0.068
	T=1000	0.096	0.097	0.095	0.100	0.106	0.111	0.107	0.124	0.118	0.095	0.093	0.065
							$\phi = 0.98$						DM
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda^0_2)$	T=250	0.090	0.092	0.093	0.091	0.094	0.092	0.098	0.107	0.093	0.066	0.059	0.034
	T=500	0.092	0.094	0.094	0.093	0.095	0.097	0.095	0.095	0.077	0.064	0.071	0.026
	T=1000	0.094	0.096	0.095	0.096	0.100	0.099	0.096	0.098	0.078	0.070	0.072	0.022
													CW
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda^0_2)$	T=250	0.108	0.113	0.119	0.117	0.123	0.129	0.146	0.179	0.183	0.150	0.121	0.094
	T=500	0.103	0.107	0.108	0.109	0.117	0.118	0.124	0.141	0.140	0.121	0.110	0.078
	T=1000	0.100	0.103	0.102	0.105	0.110	0.113	0.114	0.125	0.121	0.101	0.094	0.071

Table S18: **DGP2** Empirical Size of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ under conditional homoskedasticity; $\pi_0 = 0.50$

	λ_2^0	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	
												DM
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$	T=250	0.080	0.080	0.082	0.073	0.076	0.075	0.074	0.073	0.069	0.069	0.009
	T=500	0.086	0.087	0.087	0.086	0.085	0.081	0.081	0.078	0.073	0.066	0.009
	T=1000	0.084	0.089	0.090	0.091	0.091	0.087	0.092	0.085	0.082	0.076	0.011
												CW
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$	T=250	0.109	0.118	0.117	0.113	0.115	0.120	0.121	0.131	0.139	0.162	0.079
	T=500	0.102	0.108	0.108	0.109	0.111	0.109	0.112	0.113	0.114	0.129	0.070
	T=1000	0.099	0.102	0.105	0.108	0.108	0.110	0.116	0.115	0.115	0.122	0.077

Table S19: **DGP2** Empirical Size of $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$ under conditional homoskedasticity; $\pi_0 = 0.50$

	λ_2^0	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00	
													DM
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$	T=250	0.079	0.080	0.081	0.075	0.075	0.072	0.071	0.074	0.069	0.048	0.039	0.009
	T=500	0.087	0.089	0.087	0.085	0.083	0.078	0.077	0.069	0.062	0.051	0.047	0.009
	T=1000	0.086	0.090	0.088	0.089	0.087	0.085	0.084	0.076	0.065	0.051	0.061	0.011
													CW
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$	T=250	0.113	0.119	0.122	0.119	0.124	0.132	0.150	0.190	0.210	0.168	0.132	0.079
	T=500	0.106	0.108	0.112	0.112	0.116	0.117	0.132	0.146	0.163	0.142	0.113	0.070
	T=1000	0.100	0.105	0.106	0.110	0.108	0.112	0.122	0.136	0.136	0.117	0.107	0.077

S23 with Table 8 in the main text. If we consider the case of $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0 = 0.7)$ under $T = 250$ as an example, we have a power estimate of 99.1% under $\pi_0 = 0.25$ (Table 8 in main text) compared with 97.8% in Table S23.

Table S20: **DGP1** Empirical Power of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ under conditional homoskedasticity; $\pi_0 = 0.50$

β	0.00	-1.50	-1.75	-2.00	-2.25	-2.50	-3.00	-3.50
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$				$\phi_1 = 0.75$				
$\lambda_2^0 = 0.80$	0.095	0.187	0.229	0.282	0.333	0.396	0.533	0.671
$\lambda_2^0 = 0.85$	0.095	0.202	0.257	0.315	0.377	0.450	0.604	0.738
$\lambda_2^0 = 0.90$	0.100	0.232	0.301	0.377	0.451	0.536	0.687	0.810
$\lambda_2^0 = 0.95$	0.092	0.290	0.387	0.494	0.576	0.662	0.804	0.895
DM	0.018	0.298	0.390	0.480	0.547	0.625	0.747	0.841
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)_{adj}$								
$\lambda_2^0 = 0.80$	0.108	0.305	0.392	0.497	0.590	0.683	0.838	0.934
$\lambda_2^0 = 0.85$	0.108	0.342	0.447	0.555	0.658	0.750	0.887	0.958
$\lambda_2^0 = 0.90$	0.117	0.406	0.520	0.639	0.739	0.827	0.930	0.979
$\lambda_2^0 = 0.95$	0.116	0.513	0.648	0.764	0.847	0.906	0.971	0.994
CW	0.061	0.655	0.778	0.859	0.913	0.950	0.982	0.996
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$				$\phi_1 = 0.95$				
$\lambda_2^0 = 0.80$	0.095	0.548	0.659	0.759	0.843	0.897	0.960	0.983
$\lambda_2^0 = 0.85$	0.094	0.602	0.713	0.804	0.880	0.922	0.974	0.989
$\lambda_2^0 = 0.90$	0.093	0.667	0.779	0.858	0.914	0.945	0.982	0.992
$\lambda_2^0 = 0.95$	0.085	0.761	0.856	0.916	0.951	0.970	0.991	0.996
DM	0.022	0.728	0.813	0.876	0.917	0.941	0.977	0.988
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)_{adj}$								
$\lambda_2^0 = 0.80$	0.110	0.790	0.882	0.937	0.971	0.986	0.998	0.999
$\lambda_2^0 = 0.85$	0.109	0.829	0.913	0.954	0.981	0.993	0.999	1.000
$\lambda_2^0 = 0.90$	0.115	0.876	0.943	0.972	0.989	0.997	0.999	1.000
$\lambda_2^0 = 0.95$	0.115	0.931	0.972	0.986	0.996	0.999	1.000	1.000
CW	0.068	0.957	0.982	0.991	0.996	0.999	1.000	1.000
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$				$\phi_1 = 0.98$				
$\lambda_2^0 = 0.80$	0.093	0.789	0.869	0.923	0.951	0.971	0.991	0.997
$\lambda_2^0 = 0.85$	0.094	0.824	0.897	0.938	0.963	0.980	0.993	0.998
$\lambda_2^0 = 0.90$	0.092	0.862	0.921	0.956	0.973	0.987	0.995	0.998
$\lambda_2^0 = 0.95$	0.083	0.910	0.951	0.975	0.986	0.993	0.998	1.000
DM	0.027	0.880	0.927	0.954	0.971	0.982	0.994	0.999
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)_{adj}$								
$\lambda_2^0 = 0.80$	0.113	0.925	0.966	0.987	0.994	0.997	1.000	1.000
$\lambda_2^0 = 0.85$	0.112	0.944	0.975	0.991	0.996	0.998	1.000	1.000
$\lambda_2^0 = 0.90$	0.116	0.960	0.984	0.995	0.997	0.999	1.000	1.000
$\lambda_2^0 = 0.95$	0.118	0.978	0.992	0.997	0.999	0.999	1.000	1.000
CW	0.082	0.988	0.994	0.998	0.999	1.000	1.000	1.000

Table S21: **DGP1** Empirical Power of $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$ under conditional homoskedasticity; $\pi_0 = 0.50$

β	0.00	-1.50	-1.75	-2.00	-2.25	-2.50	-3.00	-3.50
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$				$\phi_1 = 0.75$				
$\lambda_2^0 = 0.80$	0.091	0.259	0.323	0.412	0.474	0.563	0.713	0.831
$\lambda_2^0 = 0.85$	0.094	0.346	0.442	0.542	0.625	0.711	0.833	0.909
$\lambda_2^0 = 0.90$	0.084	0.415	0.524	0.627	0.699	0.783	0.876	0.939
$\lambda_2^0 = 0.95$	0.073	0.361	0.460	0.562	0.644	0.723	0.840	0.917
$\lambda_2^0 = 1$	0.072	0.282	0.362	0.449	0.530	0.619	0.749	0.857
DM	0.018	0.298	0.390	0.480	0.547	0.625	0.747	0.841
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$								
$\lambda_2^0 = 0.80$	0.112	0.439	0.550	0.673	0.760	0.835	0.941	0.982
$\lambda_2^0 = 0.85$	0.122	0.575	0.693	0.798	0.872	0.923	0.975	0.994
$\lambda_2^0 = 0.90$	0.130	0.650	0.770	0.852	0.914	0.952	0.988	0.997
$\lambda_2^0 = 0.95$	0.109	0.587	0.711	0.811	0.877	0.927	0.979	0.994
$\lambda_2^0 = 1$	0.096	0.485	0.606	0.706	0.799	0.867	0.951	0.986
CW	0.061	0.655	0.778	0.859	0.913	0.950	0.982	0.996
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$				$\phi_1 = 0.95$				
$\lambda_2^0 = 0.80$	0.099	0.690	0.790	0.861	0.917	0.946	0.983	0.991
$\lambda_2^0 = 0.85$	0.095	0.788	0.869	0.920	0.955	0.971	0.991	0.996
$\lambda_2^0 = 0.90$	0.079	0.840	0.905	0.945	0.967	0.979	0.992	0.997
$\lambda_2^0 = 0.95$	0.072	0.800	0.874	0.923	0.956	0.972	0.990	0.996
$\lambda_2^0 = 1$	0.067	0.724	0.813	0.879	0.925	0.953	0.983	0.994
DM	0.022	0.728	0.813	0.876	0.917	0.941	0.977	0.988
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$								
$\lambda_2^0 = 0.80$	0.120	0.882	0.944	0.973	0.988	0.996	0.999	1.000
$\lambda_2^0 = 0.85$	0.131	0.942	0.975	0.988	0.996	0.999	1.000	1.000
$\lambda_2^0 = 0.90$	0.127	0.959	0.985	0.992	0.997	0.999	1.000	1.000
$\lambda_2^0 = 0.95$	0.113	0.944	0.976	0.988	0.995	0.999	1.000	1.000
$\lambda_2^0 = 1$	0.097	0.908	0.952	0.979	0.991	0.996	1.000	1.000
CW	0.068	0.957	0.982	0.991	0.996	0.999	1.000	1.000
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$				$\phi_1 = 0.98$				
$\lambda_2^0 = 0.80$	0.092	0.867	0.923	0.957	0.973	0.984	0.996	0.999
$\lambda_2^0 = 0.85$	0.090	0.915	0.952	0.975	0.986	0.991	0.998	0.999
$\lambda_2^0 = 0.90$	0.077	0.936	0.965	0.983	0.989	0.994	0.998	1.000
$\lambda_2^0 = 0.95$	0.068	0.918	0.953	0.974	0.985	0.992	0.998	1.000
$\lambda_2^0 = 1$	0.068	0.882	0.928	0.960	0.977	0.986	0.995	0.999
DM	0.027	0.880	0.927	0.954	0.971	0.982	0.994	0.999
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$								
$\lambda_2^0 = 0.80$	0.124	0.961	0.985	0.996	0.998	0.999	1.000	1.000
$\lambda_2^0 = 0.85$	0.135	0.981	0.993	0.998	0.999	0.999	1.000	1.000
$\lambda_2^0 = 0.90$	0.138	0.988	0.995	0.998	0.999	1.000	1.000	1.000
$\lambda_2^0 = 0.95$	0.123	0.981	0.992	0.997	0.999	1.000	1.000	1.000
$\lambda_2^0 = 1$	0.105	0.968	0.988	0.995	0.998	0.999	1.000	1.000
CW	0.082	0.988	0.994	0.998	0.999	1.000	1.000	1.000

Table S22: **DGP2** Empirical Power of $\mathcal{S}^0_T(\lambda^0_1 = 1, \lambda^0_2)$ and $\mathcal{S}^0_{T,adj}(\lambda^0_1 = 1, \lambda^0_2)$ under conditional homoskedasticity; $\pi_0 = 0.50$

	λ^0_2	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	
												DM
$\mathcal{S}^0_T(\lambda^0_1 = 1, \lambda^0_2)$	T=250	0.488	0.525	0.573	0.618	0.671	0.716	0.773	0.830	0.883	0.937	0.796
	T=500	0.701	0.748	0.799	0.842	0.882	0.920	0.948	0.972	0.988	0.996	0.978
	T=1000	0.895	0.931	0.954	0.973	0.985	0.993	0.997	0.999	1.000	1.000	1.000
												CW
$\mathcal{S}^0_{T,adj}(\lambda^0_1 = 1, \lambda^0_2)$	T=250	0.859	0.891	0.914	0.940	0.960	0.973	0.985	0.991	0.996	0.999	0.998
	T=500	0.975	0.985	0.992	0.996	0.998	1.000	1.000	1.000	1.000	1.000	1.000
	T=1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table S23: **DGP2** Empirical Power of $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda^0_2)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda^0_2)$ under conditional homoskedasticity;$\pi_0 = 0.50$

	λ^0_2	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00	
													DM
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda^0_2)$	T=250	0.501	0.548	0.600	0.671	0.732	0.799	0.876	0.932	0.951	0.941	0.907	0.774
	T=500	0.710	0.770	0.825	0.880	0.925	0.961	0.985	0.995	0.998	0.996	0.988	0.974
	T=1000	0.900	0.938	0.966	0.982	0.994	0.999	1.000	1.000	1.000	1.000	1.000	1.000
													CW
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda^0_2)$	T=250	0.869	0.904	0.934	0.958	0.978	0.990	0.996	0.999	0.999	0.999	0.997	0.998
	T=500	0.979	0.988	0.995	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	T=1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

5. Unit-Roots and Cointegration among predictors

DGP3 consists of a predictive regression with predictors that form a cointegrated system. The larger model contains four predictors driven by a VECM while the restricted specification contains two predictors. Specifically, $y_{t+1} = \beta_0 + \sum_{j=1}^{4} \beta_j x_{jt} + u_{t+1}$ and the predictors are generated from $\Delta X_t = \Pi X_{t-1} + v_t$ with $\Pi = \alpha\beta'$, $v_t \sim N(0, I_4)$ and $u_t \sim N(0, 1)$. We take $\alpha_{4\times 1} = (0.0, 0.3, 0.0, 0.2)'$ and $\beta_{1\times 4} = (1.0\ \ -0.5\ 0.0\ 0.0)$. We note that the system of predictors has cointegrating rank equal to two and two common trends.

Our size experiments set $\beta_3 = \beta_4 = 0$ with $\{\beta_0, \beta_1, \beta_2\} = \{0.15, 0.30, -0.15\}$ (see Tables S24-S25), while for power, we fix $T = 500$ and consider a selection of departures from the null for β_3 and β_4 (see Tables S26-S27). It is again clear that both $\mathcal{S}_{T,adj}(\lambda^0_1, \lambda^0_2)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0; \lambda^0_2)$ show good size control across nearly all parameterisations. The average based statistic continues to display a certain degree of oversizeness when λ^0_2 is set at 0.90 (e.g., 15.1% for $T = 1000$). This is again intuitively explained by the fact that $(\tau_0 = 0.8; \lambda^0_2 = 0.90)$ is the configuration that maximises local power.

The main motivation behind this DGP3 was to assess whether our earlier results on the size and power properties of the proposed tests continued to hold when the system of predictors is allowed to also contain exact unit roots. Tables S24-S25) make it clear that very little appears to change under such settings, both qualitatively and quantitatively.

Table S24: **DGP3** Empirical Size of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ under conditional homoskedasticity

	λ_2^0	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	
												DM
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$	T=250	0.045	0.045	0.045	0.042	0.043	0.040	0.038	0.036	0.033	0.027	0.001
	T=500	0.059	0.057	0.060	0.060	0.057	0.056	0.056	0.051	0.046	0.040	0.001
	T=1000	0.063	0.064	0.063	0.064	0.066	0.067	0.061	0.061	0.057	0.047	0.001
												CW
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$	T=250	0.097	0.094	0.097	0.103	0.108	0.112	0.117	0.123	0.141	0.161	0.061
	T=500	0.095	0.096	0.099	0.100	0.101	0.108	0.114	0.113	0.124	0.143	0.064
	T=1000	0.089	0.089	0.092	0.094	0.098	0.101	0.103	0.106	0.109	0.129	0.063

Table S25: **DGP3** Empirical Size of $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$ under conditional homoskedasticity

	λ_2^0	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00	
													DM
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$	T=250	0.043	0.041	0.042	0.039	0.039	0.040	0.030	0.024	0.019	0.012	0.012	0.000
	T=500	0.058	0.057	0.058	0.057	0.050	0.048	0.043	0.034	0.022	0.017	0.027	0.001
	T=1000	0.072	0.072	0.071	0.069	0.068	0.056	0.054	0.038	0.029	0.033	0.038	0.001
													CW
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$	T=250	0.099	0.098	0.105	0.109	0.121	0.134	0.149	0.190	0.217	0.186	0.154	0.066
	T=500	0.095	0.098	0.102	0.107	0.110	0.118	0.139	0.160	0.180	0.159	0.135	0.063
	T=1000	0.099	0.103	0.101	0.104	0.106	0.105	0.114	0.134	0.151	0.133	0.113	0.062

Tables S26-S27) document empirical powers associated with DGP3. We continue to note that for both test statistics power converges towards one as we increase the magnitudes of (β_3, β_4). As expected from our local power analysis, $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0 \approx 0.90)$ displays the most favorable power outcomes and for magnitudes of (β_3, β_4) near zero dominates $\mathcal{S}_{T,adj}(\lambda_1^0, \lambda_2^0)$ based inferences.

Table S26: **DGP3** Empirical Power of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ under conditional homoskedasticity

β_3	0.01	0.02	0.03	0.04	0.05	0.06	0.10
β_4	0.01	0.02	0.03	0.04	0.05	0.06	0.10
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$							
$\lambda_2^0 = 0.80$	0.073	0.189	0.394	0.625	0.805	0.913	0.999
$\lambda_2^0 = 0.85$	0.074	0.212	0.449	0.686	0.851	0.939	0.999
$\lambda_2^0 = 0.90$	0.086	0.248	0.523	0.760	0.896	0.963	1.000
$\lambda_2^0 = 0.95$	0.094	0.316	0.624	0.839	0.944	0.983	1.000
DM	0.020	0.204	0.532	0.797	0.922	0.973	1.000
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)_{adj}$							
$\lambda_2^0 = 0.80$	0.187	0.446	0.710	0.886	0.961	0.989	1.000
$\lambda_2^0 = 0.85$	0.212	0.499	0.762	0.916	0.975	0.994	1.000
$\lambda_2^0 = 0.90$	0.248	0.568	0.821	0.947	0.987	0.997	1.000
$\lambda_2^0 = 0.95$	0.317	0.666	0.884	0.975	0.995	0.999	1.000
CW	0.283	0.720	0.924	0.989	0.998	1.000	1.000

Table S27: **DGP3** Empirical Power of $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$ under conditional homoskedasticity

β_3	0.01	0.02	0.03	0.04	0.05	0.06	0.10
β_4	0.01	0.02	0.03	0.04	0.05	0.06	0.10
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$							
$\lambda_2^0 = 0.80$	0.078	0.255	0.516	0.751	0.894	0.964	0.999
$\lambda_2^0 = 0.85$	0.084	0.334	0.631	0.850	0.948	0.984	1.000
$\lambda_2^0 = 0.90$	0.093	0.382	0.706	0.895	0.966	0.992	1.000
$\lambda_2^0 = 0.95$	0.075	0.350	0.666	0.864	0.955	0.988	1.000
$\lambda_2^0 = 1$	0.076	0.290	0.581	0.808	0.922	0.974	1.000
DM	0.016	0.205	0.533	0.798	0.920	0.972	1.000
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$							
$\lambda_2^0 = 0.80$	0.260	0.573	0.821	0.948	0.986	0.997	1.000
$\lambda_2^0 = 0.85$	0.339	0.682	0.897	0.975	0.995	0.999	1.000
$\lambda_2^0 = 0.90$	0.393	0.745	0.926	0.985	0.998	1.000	1.000
$\lambda_2^0 = 0.95$	0.358	0.711	0.909	0.979	0.997	0.999	1.000
$\lambda_2^0 = 1$	0.297	0.639	0.871	0.965	0.992	0.998	1.000
CW	0.281	0.712	0.929	0.987	0.999	1.000	1.000