Appendix A: Proof of Proposition 2

Let us consider $\{X_{t+1}\}_{t=\infty}^{+\infty}$ following an $ARIMA(p,d,q)$ process, where $W_{t+1} = (1-B)^d X_{t+1}$ represents a stationary and invertible $ARMA(p,q)$ process, with B a backshift operator ($B^j Z_t = Z_{t-j}$). Thus,

$$W_t - \sum_{j=1}^{p} \phi_j W_{t-j} = \delta + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i,}$$
$$\phi(B)(1-B)^d X_{t+1} = \delta + \theta(B)\varepsilon_{t+1}.$$

Following Box, Jenkins, and Reinsel (2008), let us define the generalized autoregressive operator as:

$$Q(B) \equiv \phi(B)(1-B)^d = \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)(1-B)^d,$$
$$= \left(1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_{p+d} B^{p+d}\right).$$

Hence, we could write the process X_{t+1} as follows:

$$Q(B)X_t = \delta + \theta(B)\varepsilon_t.$$

For every single forecasting horizon h, the optimal forecast satisfies:

$$X_t^f(h) = \begin{cases} \sum_{i=1}^{p+d} \varphi_i X_t^f(h-i) + \delta - \sum_{i=l}^{q} \theta_i \varepsilon_{t+h-i}, & if\ h \leq q \\ \sum_{i=1}^{p+d} \varphi_i X_t^f(h-i) + \delta, & if\ h > q \end{cases}, \qquad (A1)$$

The general solution for the homogeneous difference equation (A1) when $h > q$ is given by:

$$X_t^f(h) = \sum_{i=1}^{p} c_i(t) m_i^h + [b_0(t) + b_1(t)h + b_2(t)h^2 + \cdots + b_{d-1}(t)h^{d-1}],$$

where $c(t)$ and $b(t)$ represents adaptive coefficients, that is, coefficients that are stochastic and functions of the process at time t, and the terms m_i corresponds to the roots of the following expression:

$$m^p - \phi_1 m^{p-1} - \phi_2 m^{p-2} - \cdots - \phi_p = 0.$$

Expression (A1) is not homogeneous, so we need to add a particular solution, which is given by:

$$b_d l^d,$$

where b_d is a deterministic coefficient. Thus, the eventual or explicit forecast function is given by:

$$X_t^f(h) = \sum_{i=1}^{p} c_i(t) m_i^h + (b_0(t) + b_1(t)h + b_2(t)h^2 + \cdots \qquad (A2)$$
$$\ldots + b_{d-1}(t)h^{d-1}) + \mathbb{I}_{(\delta \neq 0)} b_d h^d; \qquad and$$
$$h > q - p - d.$$

The previous expression characterizes long horizon forecasts from $ARIMA(p,d,q)$ models. It is interesting that the moving average terms only play a role in the determination of the adaptive coefficients. Besides, stationary roots of the autoregressive operator will vanish when the forecasting horizon lengthens as they have an absolute value less than one.

Finally, the influence of unit roots determines the presence of a polynomial of order d in the forecast horizon h, in which some of the coefficients are adaptive. If the econometrician mistakenly considers that the Y_t process follows an $ARIMA(p,1,q)$ then he or she will compute the forecasts according to (A2). When $\delta = 0$, and for large values of h we will have that the MSPE is given by:

$$MSPE(h) = \mathbb{E}\left[Y_{t+h} - Y_t^f(h)\right]^2,$$
$$MSPE(h) = \mathbb{E}\left[Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h - b_0(t)\right]^2,$$
$$MSPE(h) = \mathbb{V}\left[Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h - b_0(t)\right] + \left[\mathbb{E}\left[Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h - b_0(t)\right]\right]^2,$$
$$MSPE(h) = \mathbb{V}\left[Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h\right] + \mathbb{V}[b_0(t)] - 2\mathbb{C}\left[Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h, b_0(t)\right]$$
$$+ \left[\mathbb{E}[Y_t] - \sum_{i=1}^{p} m_i^h\, \mathbb{E}\big[c_i(t) - \mathbb{E}[b_0(t)]\big]\right]^2.$$

We notice that:

$$\lim_{h\to\infty} \left[\mathbb{E}[Y_t] - \sum_{i=1}^{p} m_i^h\, \mathbb{E}[c_i(t)] - \mathbb{E}[b_0(t)]\right]^2 + \mathbb{V}[b_0(t)]$$
$$= \left[\mathbb{E}[Y_t] - \mathbb{E}[b_0(t)]\right]^2 + \mathbb{V}[b_0(t)].$$

Therefore we will place attention on the following term:

$$\mathbb{V}\left[Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h\right] - 2\mathbb{C}\left[Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h, b_0(t)\right].$$

First, notice that:

$$\mathbb{V}\left[Y_{t+h}-\sum_{i=1}^{p} c_i(t)m_i^h\right]=\mathbb{V}[Y_{t+h}]+\mathbb{V}\left[\sum_{i=1}^{p} c_i(t)\, m_i^h\right]-2\mathbb{C}\left[Y_{t+h},\sum_{i=1}^{p} c_i(t)m_i^h\right],$$

$$=\mathbb{V}[Y_t]+\mathbb{V}\left[\sum_{i=1}^{p} c_i(t)\, m_i^h\right]-2\mathbb{C}\sum_{i=1}^{p}\left[Y_{t+h},c_i(t)m_i^h\right],$$

$$=\mathbb{V}[Y_t]+\sum_{i=1}^{p}\left|m_i^h\right|^2\mathbb{V}[c_i(t)]+\sum_{i=1}^{p}\sum_{j\neq i}\mathbb{C}\left[c_i(t)m_i^h,c_j(t)m_j^h\right]$$

$$-2\textstyle\sum_{i=1}^{p}\mathbb{C}\left[Y_{t+h},c_i(t)m_i^h\right].$$

Therefore,

$$\mathbb{V}\left[Y_{t+h}-\sum_{i=1}^{p} c_i(t)m_i^h\right]\leq\mathbb{V}[Y_t]+\sum_{i=1}^{p}\left|m_i^h\right|^2\mathbb{V}[c_i(t)]+$$

$$+2\sum_{i=1}^{p}\sum_{j<i}^{p}\left|\mathbb{C}\left[c_i(t)m_i^h,c_j(t)m_j^h\right]\right|+2\sum_{i=1}^{p}\left|\mathbb{C}\left[Y_{t+h},c_i(t)m_i^h\right]\right|,$$

$$\leq\mathbb{V}[Y_t]+\sum_{i=1}^{p}\left|m_i^h\right|^2\mathbb{V}[c_i(t)]+2\sum_{i=1}^{p}\sum_{j<i}^{p}\left|m_i^h\overline{m_j^h}\right|\mathbb{C}\left[c_i(t),c_j(t)\right]+$$

$$+2\sum_{i=1}^{p}\left|\overline{m_i^h}\right|\left|\mathbb{C}[Y_{t+h},c_i(t)]\right|,$$

$$\leq\mathbb{V}[Y_t]+\sum_{i=1}^{p}\left|m_i^h\right|^2\mathbb{V}[c_i(t)]+2\sum_{i=1}^{p}\sum_{j<i}^{p}\left|m_i^h\right|\left|m_j^h\right|\sqrt{\mathbb{V}[c_i(t)]\mathbb{V}\left[c_j(t)\right]}+$$

$$+2\sum_{i=1}^{p}\left|m_i^h\right|\sqrt{\mathbb{V}\left[c_i(t)\mathbb{V}[Y_t]\right]},$$

$$=\mathbb{V}[Y_t]+\sum_{i=1}^{p}|m_i|^{2h}\mathbb{V}[c_i(t)]+2\sum_{i=1}^{p}\sum_{j<i}^{p}|m_i|^h\left|m_j\right|^h\sqrt{\mathbb{V}[c_i(t)]\mathbb{V}\left[c_j(t)\right]}+$$

$$+2\textstyle\sum_{i=1}^{p}|m_i|^h\sqrt{\mathbb{V}\left[c_i(t)\mathbb{V}[Y_t]\right]}.$$

So

$$\lim_{h\to\infty}\mathbb{V}\left[Y_{t+h}-\sum_{i=1}^{p} c_i(t)m_i^h\right]\leq\mathbb{V}[Y_t],$$

provided that the absolute value of the stationary roots have absolute value less than one and therefore:

$$\lim_{h\to\infty}|m_i|^h=0.,$$

Now,

$$\left|\mathbb{C}\left[Y_{t+h}-\sum_{i=1}^{p} c_i(t)m_i^h\,, b_0(t)\right]\right| = \left|\mathbb{C}[Y_{t+h}, b_0(t)] - \mathbb{C}\left[\sum_{i=1}^{p} c_i(t)m_i^h, b_0(t)\right]\right|,$$

$$= \left|\mathbb{C}[Y_{t+h}, b_0(t)] - \sum_{i=1}^{p} \mathbb{C}\left[c_i(t)m_i^h, b_0(t)\right]\right|,$$

$$\leq |\mathbb{C}[Y_{t+h}, b_0(t)]| + \sum_{i=1}^{p}\left|\mathbb{C}\left[c_i(t)m_i^h, b_0(t)\right]\right|,$$

$$\leq \sqrt{\mathbb{V}[Y_{t+h}]\mathbb{V}[b_0(t)]} + \sum_{i=1}^{p}\left|m_i^h\right|\sqrt{\mathbb{V}[c_i(t)]\mathbb{V}[b_0(t)]},$$

$$= \sqrt{\mathbb{V}[Y_t]\mathbb{V}[b_0(t)]} + \sum_{i=1}^{p}|m_i|^h\sqrt{\mathbb{V}[c_i(t)]\mathbb{V}[b_0(t)]},$$

$$\underset{h\to\infty}{\rightarrow} \sqrt{\mathbb{V}[Y_t]\mathbb{V}[b_0(t)]}.$$

Finally,

$$\lim_{h\to\infty} MSPE(h) \leq \mathbb{V}[Y_t] + \sqrt{\mathbb{V}[Y_t]\mathbb{V}[b_0(t)]} + \left[\mathbb{E}[Y_t] - \mathbb{E}[b_0(t)]\right]^2 + \mathbb{V}[b_0(t)],$$

implying that the $MSPE(h)$ is a bounded sequence as the terms on the right hand side do not depend on h.

Appendix B: Proof of Proposition 3

Let Y_t be a stationary process as in Proposition 1. Let us also consider a white noise process $\{\varepsilon_{t+1}\}_{t=-\infty}^{+\infty}$ with variance σ_ε^2. Suppose that the econometrician mistakenly thinks that Y_t follows a driftless $ARIMA(p,1,q)$ process

$$\phi(B)(1-B)Y_t = \theta(B)\varepsilon_t,$$

where

$$\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i,$$
$$\theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i,$$
$$B^i X_t = X_{t-i}.$$

This process can also be written as:

$$Q(B)Y_t = \theta(B)\varepsilon_t,$$
$$Q(B) \equiv \phi(B)(1-B) = \left(1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_{p+1}B^{p+1}\right).$$

In other words, the econometrician thinks that:

$$Y_{t+h} = \sum_{i=1}^{p+1} \varphi_i Y_{t+h-i} + \varepsilon_{t+h} - \sum_{i=1}^{q} \theta_i \varepsilon_{t+h-i},$$

it follows that:

$$Q(1) \equiv \phi(1)(1-1) = 0 = (1 - \varphi_1 - \varphi_2 - \ldots - \varphi_{p+1}),$$

thus:

$$\sum_{i=1}^{p+1} \varphi_i = 1.$$

According to his or her belief, the optimal linear h-step-ahead forecast $Y_t^f(h)$ is given by:

$$Y_t^f(h) = \sum_{i=1}^{p+1} \varphi_i Y_t^f(h-i) + \varepsilon_t^f(h) - \sum_{i=1}^{q} \theta_i \varepsilon_t^f(h-i),$$

where:

$$Y_t^f(h) = Y_{t+h} \quad if \quad h \leq 0,$$
$$\varepsilon_t^f(h) = \begin{cases} 0 & if \quad h > 0 \\ \varepsilon_{t+h} & if \quad h \leq 0. \end{cases}$$

We notice that:

$$\mathbb{E}\left[\varepsilon_t^f(h)\right] = 0 \quad for\ all\ h,$$

therefore,

$$\mathbb{E}\left[Y_t^f(h)\right] = \sum_{i=1}^{p+1} \varphi_i \mathbb{E}\left[Y_t^f(h-i)\right].$$

Let us consider the following notation:

$$\mu_h \equiv \mathbb{E}\left[Y_t^f(h)\right].$$

With this notation we face the problem of solving the following equation in differences:

$$\mu_h = \sum_{i=1}^{p+1} \varphi_i \mu_{h-i},$$

$$\mu_h = \mu \equiv \mathbb{E}[Y_t] \quad if \quad h \leq 0.$$

Let us use mathematical induction for the vector:

$$S(h) \equiv \{\mu_1, \mu_2, \mu_3, \dots, \mu_h\}_{1\times h}.$$

We wish to prove the following statement:

$$S(h) \equiv \{\mu, \dots, \mu\}_{1\times h} \quad for\ all\ h \geq 1,$$

for $h = 1$ we need to prove that

$$S(1) \equiv \{\mu_1\} = \{\mu\}.$$

So, we simply need to prove that:

$$\mu_1 = \mu.$$

To that end we write:

$$\mu_1 = \sum_{i=1}^{p+1} \varphi_i \mu_{1-i} = [\varphi_1 \mu_0] + \cdots + [\varphi_p \mu_{1-p}] + [\varphi_{p+1} \mu_{-p}] = \mu \sum_{i=1}^{p+1} \varphi_i = \mu.$$

We could also prove the proposition for $h = 2$. In this case we simply need to prove that:

$$S(2) \equiv \{\mu_1, \mu_2\} = \{\mu, \mu\},$$

which is equivalent to prove that:

$$\mu_1 = \mu \quad and \quad \mu_2 = \mu.$$

It turns out that we already know that $\mu_1 = \mu$. Therefore, we only need to prove that $\mu_2 = \mu$. To that end consider:

$$\mu_2 = \sum_{i=1}^{p+1} \varphi_i \mu_{2-i} = [\varphi_1 \mu_1] + \sum_{i=2}^{p+1} \varphi_i \mu_{2-i},$$
$$\mu_2 = [\varphi_1 \mu_1] + [\varphi_2 \mu_0] + \cdots + [\varphi_p \mu_{2-p}] + [\varphi_{p+1} \mu_{1-p}],$$
$$\mu_2 = \mu \sum_{i=1}^{p+1} \varphi_i = \mu.$$

Using the mathematical induction principle, let us assume now that for $\bar{h} > 0$ we know that:

$$S(\bar{h}) = \{\mu, \dots, \mu\}_{1\times\bar{h}}.$$

We will prove then that:

$$S(\bar{h}+1) = \{\mu, \dots, \mu\}_{1\times(1+\bar{h})}.$$

$$\mu_{\bar{h}+1} = \sum_{i=1}^{p+1} \varphi_i \mu_{\bar{h}+1-i} = [\varphi_1 \mu_{\bar{h}}] + \cdots + [\varphi_p \mu_{\bar{h}+1-p}] + [\varphi_{p+1} \mu_{\bar{h}-p}].$$

Nevertheless, under the induction hypothesis we know that:

$$\mu_j = \mu \; for \; all \; j \leq \bar{h}.$$

Therefore,

$$\mu_{\bar{h}+1} = \sum_{i=1}^{p+1} \varphi_i \mu_{\bar{h}+1-i} = [\varphi_1 \mu_{\bar{h}}] + [\varphi_2 \mu_{\bar{h}-1}] + \cdots + [\varphi_p \mu_{\bar{h}+1-p}] + [\varphi_{p+1} \mu_{\bar{h}-p}],$$
$$= \mu \sum_{i=1}^{p+1} \varphi_i = \mu.$$

In other words, we have proved that:

$$\mu_j = \mu \; for \; all \; j \in \mathbb{Z}.$$

This means that forecasts coming from the incorrect $ARIMA(p, 1, q)$ specification are unbiased because:

$$\mathbb{E}[Y_{t+h} - Y_t^f(h)] = \mu - \mu_h = \mu - \mu = 0,$$

and the proof is complete.

Appendix C: Descriptive statistics of the series

Table C1: Descriptive statistics - three samples (*)

	Mean	St. dev.	Max.	Min.	Sample
	Longest estimation sample				
Canada	2,00	1,70	6,90	-0,20	1999.10-1999.1
Sweden	3,00	3,40	12,60	-1,20	1999.10-1999.1
Switzerland	2,28	2,08	6,60	-0,20	1999.10-1999.1
United Kingdom	3,27	2,14	8,50	1,30	1999.10-1999.1
United States	2,90	1,10	6,40	1,40	1999.10-1999.1
	Evaluation sample				
Canada	2,10	0,90	4,70	-0,91	1999.2-2011.12
Sweden	1,50	1,20	4,40	-1,60	1999.2-2011.12
Switzerland	0,87	0,78	3,10	-1,20	1999.2-2011.12
United Kingdom	2,12	1,13	5,20	0,50	1999.2-2011.12
United States	2,50	1,30	5,50	-2,00	1999.2-2011.12
	Full sample				
Canada	2,10	1,30	6,90	-0,90	1990.10-2011.12
Sweden	2,10	2,40	12,60	-1,60	1990.10-2011.12
Switzerland	1,42	1,59	6,60	-1,20	1990.10-2011.12
United Kingdom	2,57	1,70	2,10	0,50	1990.10-2011.12
United States	2,70	1,20	6,40	-2,00	1990.10-2011.12

(*) Source: Authors' elaboration.

Appendix D: A decision rule

As mentioned in the main body of the paper, Tables 1, 4, and 5 show, for different DGPs, a linkage between sample size, persistence, forecasting horizon and the relative accuracy of driftless unit-root-based forecasts. These tables suggest that given a DGP, a sample size R and a forecasting horizon h, there is an invertible function between the persistence of the process ρ and the relative performance of driftless unit-root-based forecasts against forecasts coming from the true DGP with estimated parameters.

We can approximate this function via simulations to find the level of persistence $\overline{\rho(R,h)}$ above which driftless unit-root-based forecasts display higher accuracy than forecasts coming from the true DGP.[1] Table D1 shows these persistence thresholds calculated using 5000 replications of the process defined by (4). In this case we compute the thresholds considering that driftless unit-root-based forecasts are constructed as linear optimal forecasts from model (5).

In a real life application, however, the true parameter ρ is unobservable and therefore we cannot design a decision rule based on $\overline{\rho(R,h)}$. Fortunately, our simulations also suggest the existence of an invertible function between the average estimated level of persistence $\hat{\rho}$ and the relative performance of driftless unit-root-based forecasts against forecasts coming from the true DGP with estimated parameters.[2] Therefore, it is also possible via simulations to find two observable thresholds coming from the average and median of the estimates of the persistence parameter.

[1] Turner (2004) proceeds with simulations as well, in a different context, to find unobservable persistence thresholds for $AR(p)$ processes.

[2] A similar relationship seems to hold true between the median of the estimated level of persistence $\hat{\rho}$ and the relative performance of driftless unit-root based forecasts against forecasts coming from the true DGP with estimated parameters.

We denote these observable thresholds by $(\overline{\rho(\widehat{R,h})})$ and $(\overline{\rho(\widehat{R,h})_m})$ respectively. In Table D1 we also report these observable thresholds.

Table D1: Persistence threshold above which unit-root-based forecasts display lower MSPE.

SARIMA Specification (*)

	Unobservable threshold ρ			Observable threshold ρ			Observable threshold ρ		
R:	50	100	200	50	100	200	50	100	200
h=1	0,73	0,90	0,96	0,66	0,85	0,93	0,68	0,86	0,93
h=2	0,77	0,91	0,96	0,69	0,86	0,93	0,71	0,87	0,93
h=3	0,79	0,91	0,96	0,71	0,86	0,93	0,73	0,87	0,93
h=4	0,80	0,92	0,96	0,72	0,86	0,93	0,74	0,87	0,93
h=5	0,81	0,92	0,97	0,72	0,86	0,94	0,74	0,87	0,94
h=6	0,82	0,93	0,97	0,74	0,87	0,94	0,76	0,88	0,94
h=12	0,83	0,94	0,97	0,75	0,88	0,94	0,77	0,89	0,94
h=24	0,74	0,93	0,98	0,67	0,87	0,94	0,69	0,88	0,95
h=36	0,70	0,93	0,97	0,64	0,87	0,94	0,66	0,88	0,94

(*) Source: Authors' elaboration.

From Table D1 we clearly see that our thresholds are increasing with the sample size R. They also look very stable across horizons when the sample size is larger ($R = 200$, for instance).

Using one of these observable thresholds we can define the following simple operational rule. Given a sample size $R \in \{50;100;200\}$ we can estimate the model in (4), get an estimate of the persistence parameter $\hat{\rho}$ and generate a h-step ahead forecast according to the following rule:

$$Y_R^f(h) = \begin{cases} Y_{UR,R}^f(h) & if \quad \hat{\rho} > \overline{\rho(\widehat{R,h})_m} \\ Y_{S,R}^f(h) & otherwise \end{cases}.$$

where $Y_R^f(h)$ represents the forecast for Y_{R+h} made with information known until time $t = R$, $Y_{S,R}^f(h)$ represents the optimal linear h-step ahead forecast based on model (4) and $Y_{UR,R}^f(h)$ represents the optimal linear h-step ahead forecast based on model (5).

To evaluate the effectiveness of our rule, we generate again 5000 independent replications of size $T = 136$ of the process (4) for different values of the persistence parameter ρ. Then, in each replication of size $T = 136$, we considered the first $R = 100$ observations to construct three different h-step ahead forecasts with $h \in \{1\text{-}6;12;24;36\}$: the optimal linear forecast assuming the data is generated by process (4), the optimal linear forecast assuming the model is generated according to model (5) and the forecast constructed according to our rule in (10).

Table D2 shows the ratio of MSPE between forecasts generated according to our rule, and forecasts generated assuming stationarity (under the columns labelled with "S" from stationary) and the ratio of MSPE between forecasts generated according our rule and forecasts generated assuming the existence of a unit root (under the columns labelled with "UR" standing for unit root).

Table D2: Ratio of MSPE between the rule in (10) and forecasts assuming either

stationarity of the presence of a unit root (*)

	S	UR	S	UR	S	UR	S	UR	S	UR	S	UR	S	UR
ρ:	0,50		0,75		0,80		0,85		0,90		0,95		0,99	
	R=100													
h=1	1,00	0,81	1,00	0,94	1,01	0,97	1,02	0,99	1,00	1,00	0,99	1,02	0,97	1,02
h=2	1,00	0,68	1,00	0,87	1,01	0,92	1,02	0,97	1,01	1,01	0,97	1,03	0,95	1,04
h=3	1,00	0,60	1,01	0,81	1,02	0,89	1,03	0,94	1,03	1,01	0,96	1,03	0,92	1,05
h=4	1,00	0,57	1,01	0,76	1,02	0,85	1,04	0,92	1,03	1,01	0,96	1,04	0,90	1,06
h=5	1,00	0,55	1,01	0,72	1,02	0,81	1,06	0,89	1,04	1,00	0,96	1,04	0,88	1,06
h=6	1,00	0,54	1,00	0,70	1,02	0,77	1,04	0,86	1,04	0,99	0,97	1,05	0,88	1,07
h=12	1,00	0,49	1,00	0,59	1,02	0,67	1,06	0,76	1,06	0,92	0,94	1,03	0,81	1,07
h=24	1,00	0,96	1,00	0,97	1,01	0,96	1,01	0,96	1,02	0,97	0,93	0,99	0,76	1,01
h=36	1,00	0,97	1,00	0,97	1,00	0,96	1,00	0,97	1,01	0,99	0,92	1,00	0,63	1,01

(*) Figures below unity favor the rule (10). "S" stands for stationarity and "UR" stand for unit root.

Source: Authors' elaboration.

Results in Table D2 are interesting because they indicate that our extremely simple rule does a fairly good job in the detection of the forecast that should be used. For instance, when the process displays low persistence ($\rho \leq 0.975$) the rule generates forecasts with barely the same MSPE than the rule that assumes stationarity permanently. Furthermore, it also outperforms driftless unit-root-based forecasts by far. When the DGP displays high persistence ($\rho \geq 0.95$) our simple rule outperforms the "stationarity rule" but it is beaten if the forecasts are built under the assumption of a unit-root in the DGP.

It is important to point out that even when our rule is beaten, it is not outperformed by far. Table D2 indicates that in the worst case our simple rule is outperformed by a 7%. However, in the best scenario, our rule outperforms its competitors by reductions in MSPE of more than 50%. Even when the DGP is persistent, gains up to 37% are reported. Overall, taking the average over the whole Table D2, our rule outperforms the fixed rules by 6%. Let us recall that we have introduced one of the simplest rules that can be designed. It might be interesting to analyze in further research refinements in this direction.