On the use of flexible rolling-window estimation and forecasting with evidence from the great recession to the COVID-19 pandemic.

M. Artemova[1,2], F. Blasques[1,2]*, S.J. Koopman[1,2]†, Z. Zhang[3]

[1] *Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands*

[2] *Tinbergen Institute, 1082 MS Amsterdam, The Netherlands*

[3] *School of Economics, Shanghai 200444, Shanghai University, China*

January 2025

Abstract

We propose a flexible rolling window estimation procedure which improves forecasting accuracy of misspecified linear autoregressive models. The method assigns different weights to data points in the observed sample which can be useful in the presence of data generating processes featuring structural breaks, complex nonlinearities, or other time-varying properties which cannot be easily captured by model design. We show how the window can be regularized by means of cross-validation. In a set of Monte Carlo experiments we reveal that the estimation method can significantly improve the forecasting accuracy of autoregressive models. In an empirical study, we achieve higher forecasting accuracy for U.S. Industrial Production during the great recession by giving more weight to observations from past recessions. Similar findings are found for other macroeconomic time series and for the 2008-2009 global financial crisis and the COVID-19 recession in 2020.

Keywords: rolling windows, autoregressive models, structural breaks.
JEL classification codes: C10, C22, C32, C51.

*Francisco Blasques acknowledges the Dutch Research Council (NWO) for their financial support with grant Vidi.195.099.

†Siem Jan Koopman acknowledges the support from Aarhus University, Denmark, and Danish National Research Foundation (DNRF78).

1 Introduction

Linear autoregressive models have been used for the analysis and forecasting of time series observations in the vast majority of empirical studies in economics and finance of the last five decades. The flexibility of the model specification allows these models to describe relatively complex dynamics in a remarkably simple manner. From short-run temporal dependence with seasonal dynamics to long-run persistency generated by stochastic trends, the basic autoregressive model is often able to provide a better in-sample fit and a more accurate out-of-sample forecast, when compared to more intricate and complex dynamic models. Given the simplistic structure of linear reduced-form models, and the complexity of data generating processes for large market economies, it is likely that these models may suffer from some form of model misspecification. A typical way in which model misspecification can manifest itself is parameter instability. There is substantial evidence of parameter instability in economic and financial empirical studies and applications; see Stock and Watson (1996, 2007) in macroeconomic forecasting, and Wolff (1987), Schinasi and Swamy (1989) and Goyal and Welch (2003) in financial forecasting. As also discussed by Inoue et al. (2017), parameter instability is widely recognized as a crucial issue that significantly hampers the forecasting performance of econometric models; see, for example, Stock and Watson (1996), Clements and Hendry (1998), Goyal and Welch (2003), Koop and Potter (2004), Paye and Timmermann (2006), Giacomini and Rossi (2009), and Rossi (2013). Several methods have been proposed to improve forecasting performance in the presence of parameter instability. Rolling-window estimation is among the most popular with important applications in finance, see Goyal and Welch (2003), in macroeconomics, see Swanson (1998), and in exchange rate forecasting, see Molodtsova and Papell (2009), among others. Using a window of observations for estimation can be regarded as a basic weighting scheme for the observations, whereby recent observations are assigned weight one

by the estimation loss function, and observations far in the past are assigned weight zero and are effectively discarded from estimation.

We propose a new flexible rolling-window estimation procedure for the unknown parameters of the autoregressive distributed lag (ADL) model. We allow for an optimal combination of multiple windows which possibly feature different lengths, different start and end points in the data set, as well as for different importance being assigned to different data points within a given window. Overall, our procedure finds features of the flexible window such that the forecasting performance of the misspecified linear autoregressive model is optimized. We refer to this procedure as *flexible rolling-window* (FRW) estimation.

The FRW procedure can be applied generally for the analysis, modeling, and forecasting of economic and financial time series. The features of the rolling windows are regularized by cross-validation. This regularization procedure automatically identifies the important data points in the observed data set and adjusts the FRW features to deliver optimal out-of-sample forecasting accuracy for the ADL model. Further, we show that the regularization by cross-validation ensures that the FRW estimator converges to the classical MLE when the ADL model is correctly specified. In this special case, the FRW estimator converges to the 'true' time-invariant parameter vector, which is optimal for forecasting. At the same time, if the ADL model is misspecified, then the cross-validation procedure will establish FRW features that are optimal for forecasting. In this case, the FRW provides time-varying parameter estimates for the ADL model which significantly improve the forecasting accuracy of the classical MLE method.

Our FRW estimator is closely related to the estimator introduced in Oh and Patton (2024), which exploits information from a state variable to improve forecasts from imperfect models. Oh and Patton (2024) find significant performance gains in applications by leveraging a state variable which introduces weights in the loss function. They optimize hyperparameters over a grid, which can sometimes be challenging when the dimensionality

of the hyperparameter space exceeds one. While Oh and Patton (2024) focus mostly on financial applications, Dendramis et al. (2020), who proposed a novel method for forecasting that exploits similarities, document small gains over the benchmark AR(1) model in macroeconomic applications. In our approach, we simultaneously estimate the parameters of the ADL model and the features of the rolling windows by means of the profile regularization criterion, without explicit connection to a state variable or a grid search for continuous hyperparameters. Furthermore, we provide theoretical justifications for the estimation procedure and discuss forecast comparison tests in applications. Our method, which is not limited to forecasting based on similarity as in Dendramis et al. (2020), provides substantial and often significant gains over linear counterparts in macroeconomic applications.

This paper is organized as follows. Section 2 summarizes the basic concepts of rolling windows and establishes its relation to FRW. Section 3 introduces the FRW estimation procedure and its regularization by cross-validation. Section 4 investigates the theoretical aspects of FRW. In particular, we show its asymptotic equivalence to MLE, provide conditions for optimal forecasting performance, and show the asymptotic validity of the Giacomini-White test for the comparison of forecasting accuracy. For a selection of empirical illustrations in Section 6, we show that the optimal weight functions can take intuitive forms that would be difficult to obtain from time-varying parameter models. Section 7 concludes and discusses possible directions for future research.

2 ADL model and flexible rolling windows

2.1 Classical rolling windows

We consider the linear autoregressive distributed lag (ADL) model in our treatment of FRW estimation. The ADL(p, q) model assumes that the time series $\{Y_t\}_{t\in\mathbb{Z}}$ is generated

by the dynamic stochastic process

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + ... + \alpha_p Y_{t-p} + \beta_1 X_{t-1} + ... + \beta_q X_{t-q} + \epsilon_t, \qquad t \in \mathbb{Z}, \tag{1}$$

where $\{X_t\}_{t\in\mathbb{Z}}$ is an exogenous stochastic process, $p \in \mathbb{N}_+$ and $q \in \mathbb{N}_+$ are the lag-orders of Y_t and X_t, respectively, $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of iid innovations with density $p_\epsilon(\boldsymbol{\lambda})$ indexed by the vector of parameters $\boldsymbol{\lambda}$. For simplicity, we collect all parameters in the vector $\boldsymbol{\theta} := (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})$ where $\boldsymbol{\alpha} := (\alpha_0, \ldots, \alpha_p)$ and $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_q)$.

The parameter vector $\boldsymbol{\theta}_0 \in \Theta$ is easy to estimate via the exact maximum likelihood method or via regression (based on the conditional likelihood function). The ML estimator is thus given by,

$$\hat{\boldsymbol{\theta}}_T := \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{t=s}^{T} \ell_t(\boldsymbol{\theta}), \tag{2}$$

where $\ell_t(\boldsymbol{\theta})$ denotes the logarithm of the conditional density of Y_t given $P_{t-1}, P_{t-2}, \ldots, P_1$ with $P_t := (Y_t, X_t')'$, $s := 1 + \max(p, q)$ and $e_t(\boldsymbol{\alpha}, \boldsymbol{\beta})$ denotes the residual term

$$e_t(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) := Y_t - \hat{\alpha}_0 - \hat{\alpha}_1 Y_{t-1} - \ldots - \hat{\alpha}_p Y_{t-p} - \hat{\beta}_1 X_{t-1} - \ldots \hat{\beta}_q X_{t-q}, \qquad t = s, \ldots, T.$$

Once a point estimate $\hat{\boldsymbol{\theta}}_T$ is found, the autocorrelation function, the forecast function, including forecast confidence bounds, and the impulse response function for the ADL(p, q) model can easily be computed.

It is well known that rolling-window estimation procedures can substantially improve the forecasting performance of ADL models when changes in social, economic and financial systems render past data less relevant. Examples are pervasive in macroeconomic and financial forecasting. From inflation forecasting which recognizes that inflation dynamics in the 1980s was far too different from inflation dynamics in the 2010s, see Stock and Watson (2007), to money growth forecasting where past data can be equally misleading,

see Tarassow (2019). From a parameter estimation perspective, rolling-window estimation can however be seen as an extreme example of data weighting (Giraitis et al., 2013), where recent observations receive unit weight and data points far in the past receive weight zero, being effectively excluded from estimation. The top graph of Figure 1 represents this weighting scheme as a simple diagram which assigns weight w_t to the data point y_t, over a sample of points. At time T, for a window of length R, a typical rolling-window procedure assigns binary weights of the form,

$$w_t = \mathbf{1}_{(t\in[T-R,T])} \quad \text{for } t = s, ..., T,$$

where $\mathbf{1}_{(t\in[T-R,T])}$ denotes an indicator function assigning value 1 to observations t in the interval $[T-R, T]$, and zero otherwise. With such a window, the ML estimator is then given by,

$$\hat{\boldsymbol{\theta}}_T(\mathbf{w}_T) := \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{t=s}^{T} w_t \cdot \ell_t(\boldsymbol{\theta}), \tag{3}$$

with $\mathbf{w}_T := (w_s, ..., w_T)$ collecting the weights assigned to each observation. In this context, it is clear that data points with unit weights are 'included' in the estimation procedure, while data points with zero weights are not taken into account and effectively 'excluded' from estimation.

There are however many scenarios where having such a clear window cut-off point may be difficult to justify or even undesirable. The use of flexible-rolling windows aims to fill this gap.

2.2 Flexible rolling windows

FRWs can play an important role in settings where simply 'including' and 'excluding' data points from estimation is too simplistic. The second graph in Figure 1 shows, for example, how multiple windows can be combined to give positive weight to observations from specific

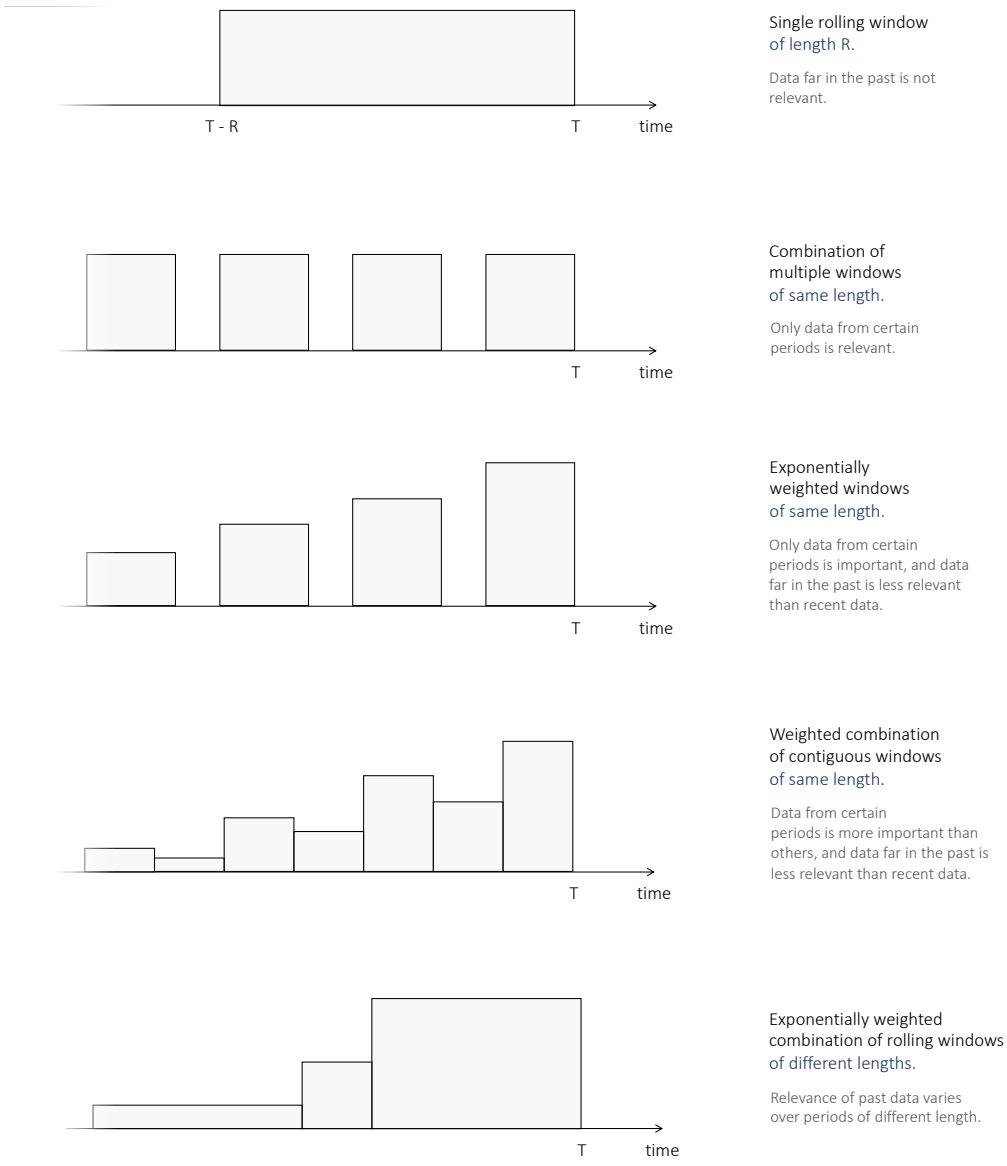


Figure 1: Examples of flexible rolling windows obtained through weighted combination of multiple windows of different length and spanning different periods.

periods only. The following weights are obtained by combining m windows with specific start and end points,

$$w_t = \sum_{i=1}^{m} \mathbf{1}_{(t \in [t_i^-, t_i^+])} \quad \text{for } t = s, ..., T.$$

Alternatively, the FRW can be used to give higher importance to data from past periods of greater historical relevance. In this way, the FRW exploits more information from specific

periods which, for example, are similar in terms of economic conditions. As we shall see in our empirical application, Industrial Production Index (IPI) forecasts during an economic recession can be significantly improved by giving higher weights to past episodes of economic recession. FRW estimation allows us to effectively learn exclusively from the specific periods of time, giving data from those periods larger weight in the estimation.

The third and fourth graphs in Figure 1 show further how different windows can be combined to ensure both that data far in the past receives less weight for estimation, but also, that data from certain periods is more relevant than others. When forecasting IPI during a recession this allows not only for data from recession periods to receive more weight, but also, for recessions far in the past to receive less weight than recent recessions. We show in our empirical study that forecasts of the U.S. Industrial Production during the global 2008–2009 recession can be significantly improved. In practice, the FRW estimation procedure can be used to give higher weights to more recent observations compared to observations far in the past. This may be desirable as various political, institutional, and technological developments change social, economic and financial systems, rendering past observations increasingly obsolete for the forecasting of economic variables. For example, as demonstrated in Boud et al. (2023), improvements in GDP nowcasts can be achieved by assigning greater weights to the periods with similar sentiment levels observed in the past. In general, the following weights for the data are obtained by combining m windows with weights $\omega_1, ..., \omega_m$,

$$w_t = \sum_{i=1}^{m} \omega_i \mathbf{1}_{(t \in [t_i^-, t_i^+])} \quad \text{for } t = s, ..., T.$$

The bottom graph in Figure 1 shows that, at any given point in time, the FRW windows can be adjusted to reflect the importance of past observations for improving forecast accuracy. In particular, it highlights that windows can have different sizes and weights, implying that short spells of relevant data can be combined with long spells of less relevant data, and

vice-versa. At the end of the day, it is only natural that multiple factors may influence data relevance. For example, forecasts for inflation during international oil crises can be improved by paying special attention to price dynamics during past oil crises. Similarly, the forecasts for unemployment during a period of strong fiscal austerity can benefit from giving more prominence to observations that originate from past episodes of strong fiscal austerity. In this way, the use of the FRW often results in a time-varying parameter ADL model whose parameter estimates are optimal for forecasting observations in a given period of interest.

3 FRW estimation and regularization

3.1 The estimator

As noted in the previous section, a simple and convenient way to formalize the influence of FRW on estimation is to assign the weights to each observation in the sample when calculating the criterion function for estimating the vector $\boldsymbol{\theta}$ of ADL(p, q) model parameters. For reasons of simplicity and computational efficiency, it will often be beneficial to parameterize the weights using a low-dimensional parameter vector $\boldsymbol{\rho}$. For example, one may include in the vector $\boldsymbol{\rho}$ features such as window length, window boundaries, or even window weights, when multiple windows are being combined. In such cases, each weight can be denoted by $w_t(\boldsymbol{\rho})$, and our FRW maximum likelihood estimator will take the form,

$$\hat{\boldsymbol{\theta}}_T(\boldsymbol{\rho}) := \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{t=s}^{T} w_t(\boldsymbol{\rho}) \cdot \ell_t(\boldsymbol{\theta}), \tag{4}$$

where $\ell_t(\boldsymbol{\theta})$ denotes the logarithm of the conditional density of Y_t given the past observations $P_{t-1}, P_{t-2}, \ldots, P_1$,

$$\ell_t(\boldsymbol{\theta}) := \log p_\epsilon\big(e_t(\boldsymbol{\alpha}, \boldsymbol{\beta}); \boldsymbol{\lambda}\big),$$

with $e_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) := Y_t - \alpha_0 - \alpha_1 Y_{t-1} - \ldots - \alpha_p Y_{t-p} - \beta_1 X_{t-1} - \ldots \beta_q X_{t-q}$ as before.

This implies that the FRW features $\boldsymbol{\rho}$ define implicitly a set of weights which, in turn, determines the parameter estimates $\hat{\boldsymbol{\theta}}_T$ for the ADL(p, q) model. In other words, we have a map $\hat{\boldsymbol{\theta}}_T : \mathcal{R} \times \Omega \to \Theta$ where $\mathcal{R} \subseteq \mathbb{R}^{d_{\boldsymbol{\rho}}}$ is the parameter space for the vector $\boldsymbol{\rho}$ of FRW features, with dimension $d_{\boldsymbol{\rho}}$, Ω is the event space of the underlying probability space of interest, and $\Theta \subseteq \mathbb{R}^{p+q+2}$ is the parameter space for the ADL(p, q) model with parameters for the intercept, lags, and innovation's variance.

3.2 Time-varying sequence of estimates

The fundamental difference between the FRW and the classical ML estimator is the introduction of weights for different observations in the log likelihood function which are implicitly defined by the rolling window(s). Observations with relatively large (small) weight will have a relatively large (small) influence in the estimation of the parameter vector $\boldsymbol{\theta}$. As such, the FRW parameter estimates will attempt to fit more accurately the dynamics of the time series at certain periods of interest.

When a dynamic model is well specified, it is well known that the classical MLE produces optimal forecasts in Kullback-Leibler divergence; see e.g. Blasques et al. (2015). However, when the data generating process changes over time, resulting in a misspecified ADL model, having a good in-sample fit does not ensure a good out-of-sample forecasting accuracy. Our Proposition 3 in Section 4 highlights that for every misspecified ADL model, there exists a non-constant sequence of parameters that improves the approximation of the ADL model to the true data-generating process, and ultimately exhibits improved forecasting accuracy. We further show that the FRW estimation procedure can help in finding such an appropriate sequence of parameters, and that the FRW estimation outperforms the classical MLE.

In general, similarly to rolling-window parameter estimates, the FRW estimates will differ from one time period to the next. For example, the parameter estimate $\hat{\boldsymbol{\theta}}_t(\mathbf{w}_t)$

obtained using the sub-sample $P_1, ..., P_t$, will typically be different from the parameter estimate $\hat{\boldsymbol{\theta}}_{t+1}(\mathbf{w}_{t+1})$ obtained using the sub-sample $P_1, ..., P_{t+1}$. For this reason, the FRW can potentially be used to construct a sequence of parameter estimates $\{\hat{\boldsymbol{\theta}}_t(\mathbf{w}_t)\}$ that describes the parameter instability in the ADL(p, q) model.

3.3 Further parameterizing features

Note that, in principle, by adopting a weighted estimation loss function, we allow for even more flexibility in the properties of rolling windows. For example, we can set exponentially decaying weights for data points within a given window by setting $\boldsymbol{\rho} = (t^-, t^+, \rho)$ where t^- and t^+ denote the lower and upper bounds of the rolling window, and ρ is a scalar parameter which determines the exponential weight decay within the interval, such that $w_t = \rho^{(t^+ - t)}$ for every $t^- \leq t \leq t^+$ and $w_t = 0$ outside of the interval $[t^-, t^+]$.

Another interesting class of weights is obtained by letting the weights depend on lagged values of Y_t or X_t and/or other variables of interest Z_t. As mentioned earlier, we show in Section 6 that forecasts of Industrial Production during the great recession, can be significantly improved by defining weights that make past recession periods more informative, and, at the same time, downweight more recent observations. In particular, we use the NBER recession indicator Z_t and define the weights of the vector $\mathbf{w}_k(\boldsymbol{\rho}, Z_t)$ as follows

$$w_t(t^+, \rho_1, \rho_2, Z_t) = \rho_1^{(t^+ - t)}(\rho_2 \cdot (1 - Z_t) + (1 - \rho_2) \cdot Z_t), \tag{5}$$

where $0 \leq \rho_1, \rho_2 \leq 1$. This specification illustrates the generality of our framework of weighing observations in the context of maximum likelihood estimation.

3.4 Regularization

It is well documented in the literature that rolling windows can provide important insights into parameter instability, be it in the form of breaks, trends, seasonality or random changes. However, without employing appropriate regularization methods for window-length, any improvements in forecast accuracy would be the result of an ad-hoc improvement in model specification. Moreover, even if the window-length is optimized, classical rolling-window estimation will provide improved forecasts only if it turns out that the exclusive use of recent data is advantageous for forecasting. In this section we propose a cross-validation method of regularization which aims to optimize not only the window-length, but all the relevant features of our FRW procedure.

In principle, the set of parameters $\boldsymbol{\rho}$ defining the FRW features can be obtained by optimizing a given criterion function q_t over a 'regularization' set as follows:

$$\hat{\boldsymbol{\rho}}_H = \arg\min_{\rho} Q_H(\boldsymbol{\rho}) = \arg\min_{\rho} \frac{1}{H} \sum_{k=T'}^{T-n} q_k(\boldsymbol{\rho}), \tag{6}$$

where we set $H = T - T' - n$. Common choices of regularization criterion q_k include squared errors, absolute errors, and likelihoods. For illustration purposes, we shall often focus on optimizing the forecast accuracy of the ADL(p, q) model in an MSE sense. In particular, we select rolling-windows' features $\boldsymbol{\rho}$ which optimize the n-step-ahead squared forecast error,

$$Q_H(\boldsymbol{\rho}) = \frac{1}{T - T' - n} \sum_{k=T'}^{T-n} \left(\hat{Y}_{k+n}\big(\hat{\boldsymbol{\theta}}_k(\boldsymbol{\rho})\big) - Y_{k+n} \right)^2, \tag{7}$$

where the FRW $\hat{\boldsymbol{\theta}}_k(\boldsymbol{\rho})$ depends on the feature vector $\boldsymbol{\rho}$ of the flexible rolling windows, and T' defines the sample point from which the forecasting accuracy begins to be measured. A large T' gives us more data to estimate the parameter vector $\boldsymbol{\theta}$ by FRW, but a small number of observations to evaluate the forecasting accuracy of the model and optimize the

weights. On the contrary, a small T' increases the uncertainty in the estimation of $\boldsymbol{\theta}$ but gives us a larger sample to determine the optimal weights. Figure 2 shows how the data $Y_1, ..., Y_T$ is split into parameter estimation and regularization parts. Once regularization is achieved and the optimal FRW features $\hat{\boldsymbol{\rho}}_H$ are found, then a final forecast can be produced under the point estimate $\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\rho}}_H)$ as illustrated in Figure 2.

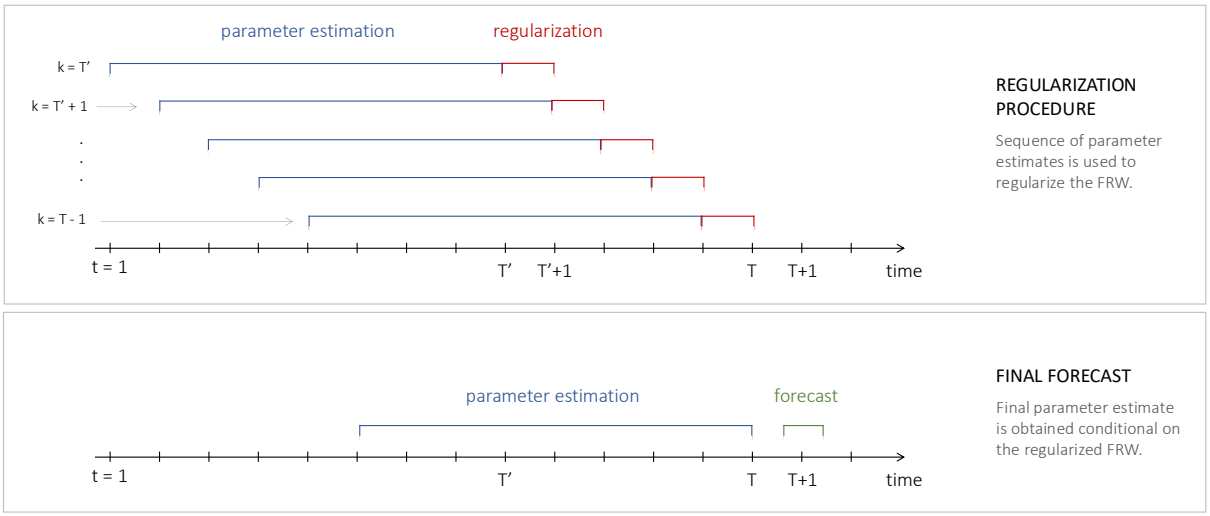


Figure 2: Data split for both the FRW regularization procedure (top), as well as the final ADL(p, q) parameter estimation and forecasting step (bottom) for one-step-ahead forecast.

3.5 Profile optimization procedure

The optimization of the FRW features in (6) relies on the relation between the features $\boldsymbol{\rho}$ and the parameter estimates $\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\rho}}_H)$ obtained as a function of the estimates of $\boldsymbol{\rho}$. The optimization would be trivially simple if, for a given data sample, the mapping $\hat{\boldsymbol{\theta}}_T : \mathbb{R}_+^T \to \Theta$ from weight vectors in $\mathbb{R}_+^T$ to point estimates in Θ were known analytically or if $\boldsymbol{\rho}$ could take only a finite number of values. In general, however, this map is analytically intractable. As a result, we need to carry out the optimization numerically.

For continuous features $\boldsymbol{\rho}$, our optimization algorithm seeks to find the optimal FRW features $\hat{\boldsymbol{\rho}}_H$ and the respective ADL(p, q) parameter estimates $\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\rho}}_H)$ simultaneously. This is achieved by optimizing over the features $\boldsymbol{\rho}$ while calculating the best ADL(p, q) parameter

$\hat{\boldsymbol{\theta}}_T(\boldsymbol{\rho})$ at each step. In this way, we essentially optimize the 'profile' regularization criterion, as we 'concentrate out' the optimal ADL(p, q) parameters at each step. This can be done since, under mild regularity conditions, each value of $\boldsymbol{\rho}$ implies a unique best parameter estimate $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\rho})$ for the vector $\boldsymbol{\theta}$. We display each step of the regularization procedure in Algorithm 1.

ALGORITHM 1: PROFILE OPTIMIZATION.

1. Set an initial $\hat{\boldsymbol{\rho}}_H^{(1)}$ and obtain ML estimates for the ADL(p,q) model $\hat{\boldsymbol{\theta}}_k^{(1)}(\hat{\boldsymbol{\rho}}_H^{(1)})$, $k = T', ..., T-n$.
2. Evaluate the regularization criterion in (6) at $\hat{\boldsymbol{\rho}}_H^{(1)}$ by producing a sequence of n-step-ahead forecasts $\hat{Y}_{k+n}(\hat{\boldsymbol{\theta}}_k^{(1)}(\hat{\boldsymbol{\rho}}_H^{(1)}))$, $k = T', ..., T-n$.
3. Update the FRW features to a new value $\hat{\boldsymbol{\rho}}_H^{(2)}$ and obtain new ML estimates for the ADL(p,q) model $\hat{\boldsymbol{\theta}}_k^{(2)}(\hat{\boldsymbol{\rho}}_H^{(2)})$, $k = T', ..., T-n$.
4. Re-evaluate the regularization criterion in (6) at $\hat{\boldsymbol{\rho}}_H^{(2)}$ by producing a new sequence of n-step-ahead forecasts $\hat{Y}_{k+n}(\hat{\boldsymbol{\theta}}_k^{(2)}(\hat{\boldsymbol{\rho}}_H^{(2)}))$, $k = T', ..., T-n$.
5. For $j \geq 3$ iterate as follows:
   - 5.1. Update $\boldsymbol{\rho}$ in the steepest descent direction to a new value $\hat{\boldsymbol{\rho}}_H^{(j)}$ and obtain ML estimates $\hat{\boldsymbol{\theta}}_k^{(j)}(\hat{\boldsymbol{\rho}}_H^{(j)})$, $k = T', ..., T-n$.
   - 5.2. Re-evaluate the regularization criterion (6) at $\hat{\boldsymbol{\rho}}_H^{(j)}$ using the new forecast sequence $\hat{Y}_{k+n}(\hat{\boldsymbol{\theta}}_k^{(j)}(\hat{\boldsymbol{\rho}}_H^{(j)}))$, $k = T', ..., T-n$.
   - 5.3. If $Q_H(\hat{\boldsymbol{\rho}}_H^{(j)}) < Q_H(\hat{\boldsymbol{\rho}}_H^{(j-1)})$:
     Repeat step 5 with $j = j+1$.
   - 5.4. If $Q_H(\hat{\boldsymbol{\rho}}_H^{(j)}) \geq Q_H(\hat{\boldsymbol{\rho}}_H^{(j-1)})$:
     Define $\hat{\boldsymbol{\rho}}_H = \hat{\boldsymbol{\rho}}_H^{(j-1)}$, collect $(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\rho}}_H), \hat{\boldsymbol{\rho}}_H)$, and stop iterating.

This simple steepest-ascent algorithm has revealed itself to be fast and stable in our Monte Carlo experiments reported in Technical Appendix B and in our empirical application in Section 6. Note that, in the first step, we start with an initial $\boldsymbol{\rho}$ and optimize the likelihood to obtain the standard MLE for $\boldsymbol{\theta}$. We further recall that $Q_H(\boldsymbol{\rho})$ denotes the n-step-ahead forecasting performance criterion chosen to optimize the FRW features $\boldsymbol{\rho}$, of which the MSFE in (7) is a special case. Since the algorithm can be initiated at the classical ML estimates, the FRW will feature non-uniform weights and differ from the MLE only when there is scope for improvement over the MLE.

4 Theoretical foundations for the FRW estimation

The estimation and regularization procedure detailed in the previous section is intuitively appealing as it leads to FRWs that improve the out-of-sample forecasting performance rather than the in-sample fit of the ADL(p, q) model. Below we provide some theoretical arguments for this procedure.

First, we analyze the MLE for FRW as a generalization of the classical MLE. In particular, we show that if the ADL(p, q) model is well specified, then the regularized FRW converges asymptotically to the MLE, and hence, uncovers the true parameter vector and minimizes forecast errors. On the other hand, we also show that, if the model is misspecified, then there exist FRW features setting non-uniform weights that improve upon the MLE parameter estimates in terms of forecasting performance. Furthermore, we show that our algorithm for finding optimal FRW features outperforms the MLE under very general conditions. Second, we give conditions under which our cross-validation procedure delivers a FRW that provides optimal forecasting performance. These results apply to a wide range of forecasting performance criteria. Third, we implement a Giacomini-White-type test that can be used to infer whether the improvements in forecasting accuracy from a change in FRW features are statistically significant or not and we analyze the validity of the asymptotic distribution of the statistic.

4.1 Time-varying ADL(p, q) representations

The FRW can be used to account for instability in the parameters of ADL(p, q) models. For example, the FRW class has recursive estimators as well as rolling-window estimators as special cases. Proposition 1 provides a simple representation result by showing that many DGPs can be written in the form of a time-varying parameter ADL(p, q) model with Gaussian innovations.

For simplicity, we restrict our attention to a DGP with stochastic contracting dynamics. In any case, the result applies to a diverse number of time series. Indeed, Y_t can depend nonlinearly on its past, as well as on a potentially very large vector V_t of variables that may include not only innovations and random breaks, but also a wide range of exogenous variables with complex dynamics and temporal dependence patterns.

PROPOSITION 1. (Time-varying ADL(p,q) representation) *Let* $\{Y_t\}_{t\in\mathbb{Z}}$ *be generated according to*

$$Y_t = \phi(Y_{t-1}, V_t) \quad , \quad t \in \mathbb{Z}, \tag{8}$$

where

(i) $\{V_t\}_{t\in\mathbb{Z}}$ *with* $V_t = (X_{t-1}, \epsilon_t)$ *is a near epoch dependent (NED) of size* $-a \equiv -2(r-1)/(r-2)$ $n_{\mathcal{V}}$*-variate stochastic sequence on a* ϕ*-mixing sequence with mixing coefficients of size* $-r/(r-1)$*;*

(ii) $\sup_t \mathbb{E}|V_t|^{4r} < \infty$ *for some* $r > 2$*;*

(iii) $\sup_{v\in\mathcal{V}} \sup_{y\in\mathcal{Y}} |\phi'_y(y,v)| < 1$*,*

(iv) $\sup_{v\in\mathcal{V}} \sup_{y\in\mathcal{Y}} |\phi'_v(y,v)| < \infty$*.*

Then the following time-varying ADL(p,q) representation holds

$$Y_t = \alpha_{0,t}+\alpha_{1,t}Y_{t-1}+...+\alpha_{p,t}Y_{t-p}+\beta_{1,t}X_{t-1}+...+\beta_{q,t}X_{t-q}+\epsilon_t\,, \quad \epsilon_t \sim N(0,\sigma^2_\epsilon)\,, \quad t \in \mathbb{Z}\,, \tag{9}$$

where $\{\alpha_{i,t}\}_{t\in\mathbb{Z}}$ *and* $\{\beta_{j,t}\}_{t\in\mathbb{Z}}$ *are NED for every* $i = 0,...,p$ *and* $j = 1,...,q$*, and furthermore* $\{Y_t\}_{t\in\mathbb{Z}}$ *is also NED of size* $-a$ *and* $\sup_t \mathbb{E}|Y_t|^{4r} < \infty$*.*

4.2 Conditions for optimal FRW forecasting

A distinct feature of the FRW estimator is the fact that it reduces to classical MLE when the weights implicitly defined on the data by rolling windows are unnecessary or undesirable. Proposition 2 highlights that for a suitably regularized FRW procedure, the weights converge in probability to unity when the model is well specified. Specifically, the cross-validation method that we propose for estimating the weights ensures that the weighted likelihood function converges in probability to the classical likelihood function as the size of the estimation sample $S := T' - \max\{p, q\}$ and cross-validation sample $H := T - T' - n$ diverge to infinity sequentially. The Monte Carlo evidence reported in Appendix B confirms that the weights remain close to unity, even in finite sample dimensions that are typical in empirical studies.

PROPOSITION 2. (Asymptotic regularization under correct specification) *Let $\mathcal{R}$ be compact and suppose the conditions of Proposition 1 hold and*

$$\phi(Y_{t-1}, X_{t-1}, \epsilon_t) = \alpha_0 + \alpha_1 Y_{t-1} + ... + \alpha_p Y_{t-p} + \beta_1 X_{t-1} + ... + \beta_q X_{t-q} + \epsilon_t \,, \ \forall \ t \in \mathbb{Z}.$$

Then the MSFE criterion in (6) ensures that $\mathbf{w}_t(\hat{\boldsymbol{\rho}}_H) \xrightarrow{p} 1 \ \forall \ t$ as $T' \to \infty$ and $H \to \infty$ sequentially, for any given forecasting horizon $n \geq 1$ and lag orders $p, q \geq 1$.

By application of Berge's Maximum Theorem, we obtain as a corollary that FRW converges in probability to the MLE as the cross-validation sample H diverges to infinity. When both the cross-validation sample H and the estimation sample T' diverge to infinity, FRW converges to the true parameter $\boldsymbol{\theta}_0 \in \Theta$, just as MLE does. The Monte Carlo evidence reported in Appendix B reveals that FRW performs well in finite samples.

COROLLARY 1. *Let the conditions of Proposition 2 hold. Then $\|\hat{\boldsymbol{\theta}}_{T'}(\hat{\boldsymbol{\rho}}_H) - \hat{\boldsymbol{\theta}}_{T'}(1)\| \xrightarrow{p} 0$ as $H \to \infty$ and $\hat{\boldsymbol{\theta}}_{T'}(\hat{\boldsymbol{\rho}}_H) \xrightarrow{p} \boldsymbol{\theta}_0$ as $T' \to \infty$ for any given forecasting horizon $n \geq 1$.*

Corollary 1 reveals that the FRW will only deliver constant parameters in large samples when the model is well specified. Under incorrect model specification, recursive or rolling-window estimators, can often improve upon full-sample estimators by allowing for time-varying parameters that better capture the dynamics of the data at any given period of time. Similarly, the FRW will be able to improve the forecasting performance of the ADL(p, q) model by allowing for time-varying parameters that can improve the out-of-sample performance of the model. The existence of such a sequence of parameters is another simple, albeit important and general, consequence of Proposition 1. Below we let $\text{MSFE}_n(\boldsymbol{\theta})$ denote the n-step ahead mean squared error achieved by the ADL(p, q) model under some parameter vector $\boldsymbol{\theta} \in \Theta$,

$$\text{MSFE}_n(\boldsymbol{\theta}) = \mathbb{E}_t\Big(Y_{t+n} - \hat{Y}_{t+n}(\boldsymbol{\theta})\Big)^2.$$

Corollary 2. *Let the conditions of Proposition 1 hold, and suppose that*

$$\phi(Y_{t-1}, V_t) \neq \alpha_0 + \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \beta_1 X_{t-1} + \cdots + \beta_q X_{t-q} + \epsilon_t\,, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2)\,,$$

for every $\boldsymbol{\theta} \in \Theta$ and any $t \in \mathbb{Z}^\star \subseteq \mathbb{Z}$. Then there exists a non-constant sequence $\{\boldsymbol{\theta}_t\}_{t\in\mathbb{Z}}$ of points in Θ such that $\text{MSFE}_n(\boldsymbol{\theta}_t) \leq \text{MSFE}_n(\boldsymbol{\theta})$ almost surely for any t and any given $\boldsymbol{\theta} \in \Theta$ and $n \geq 1$.

Corollary 2 highlights that time-varying parameters can improve the forecasting of the ADL(p, q) when the model provides a simplistic representation of the data. Proposition 3 below focuses on the properties of the FRW algorithm proposed in the previous section. First, it highlights that the algorithm is designed to ensure that the FRW outperforms (or is at least as good as) the MLE in terms of the forecasting accuracy of the ADL(p, q) model in the cross-validation sample. Furthermore, Proposition 3 shows that under appropriate

regularity conditions, the FRW algorithm will actually uncover the weights that optimize the forecasting performance of the ADL(p, q) model in the cross-validation sample. We let $\hat{\boldsymbol{\rho}}_H^{(j)}$ denote the j-th iteration weights and we let $Q_H(\boldsymbol{\rho})$ denote the mean squared error in the cross-validation sample obtained under $\boldsymbol{\rho}$. The Monte Carlo evidence reported in Appendix B and the empirical evidence in Section 6 reveal that the FRW is indeed capable of significantly improving the forecasting performance of the ADL(p, q) model.

PROPOSITION 3. *For any given realized sample* $\{y_t\}_{t=1}^T$*, Algorithm 1 ensures that*

$$Q_H(\hat{\boldsymbol{\rho}}_H^{(j+1)}) \leq Q_H(\hat{\boldsymbol{\rho}}_H^{(j)}), \quad \forall\ j \geq 1,$$

and hence the FRW outperforms the MLE under the Q_H *criterion.*

If furthermore it holds that

$$\sup_j \sup_{\boldsymbol{\rho}} \left|\partial \hat{\boldsymbol{\theta}}_k^{(j)}(\boldsymbol{\rho})/\partial \boldsymbol{\rho}\right| < 1, \quad \textit{and} \quad \sup_j \sup_{\boldsymbol{\theta}} \left|\partial \hat{\boldsymbol{\rho}}_H^{(j)}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}\right| < 1.$$

Then, $\hat{\boldsymbol{\rho}}_H^{(j)} \to \boldsymbol{\rho}^*$ *and* $\hat{\boldsymbol{\theta}}_k(\hat{\boldsymbol{\rho}}_H^{(j)}) \to \boldsymbol{\theta}_k^*$ *as* $j \to \infty$*, for any given* $n \geq 1$*.*

Two main conditions of Proposition 3 ensure the contraction of the maps $\hat{\boldsymbol{\theta}} : \boldsymbol{\rho} \mapsto \boldsymbol{\theta}$ and $\hat{\boldsymbol{\rho}} : \boldsymbol{\theta} \mapsto \boldsymbol{\rho}$. Since these maps are often not known analytically, the contracting behavior can only be verified numerically. This can be achieved by optimizing the derivatives stated above and ensuring that the maximum is less than one.

4.3 Test for forecast precision improvement

The result established in Proposition 3 is important, but it ensures only that the FRW improves the finite sample MSFE. In other words, the algorithm discussed in Section 3.3 delivers FRW features that optimize the forecasting performance within the cross-validation sample. However, due to sampling error, it is impossible to ensure that the true forecasting

performance has improved from the MLE to the FRW with some weight matrix $\mathbf{W}$. To assess whether the improvement in forecasting performance is statistically significant or not, we will utilize a simple Diebold-Mariano (DM) test statistic. However, the FRW estimator collapses to the MLE when the methods have equal forecasting accuracy; the methods are 'nested' under the null hypothesis. Hence, we will instead consider a limited memory estimator, as proposed in Giacomini and White (2006), where the parameters of the ADL model are estimated using (maximum) estimation window size $m < \bar{m} < \infty$. This choice justifies the use of the DM test statistics in the context of nested models, unlike the DM test proposed by Diebold and Mariano (1995). Proposition 4 highlights the validity of the asymptotic distribution derived by Giacomini and White (2006) when the null hypothesis compares the MLE against an alternative FRW. We also generalize the original Giacomini and White (GW) test to near epoch dependent (NED) processes, rather than mixing processes. This allows for the consideration of fairly general DGP processes in (8).

Below, we let MSFE($\mathbf{W}$) denote the MSFE achieved by the ADL model under the FRW with weights $\mathbf{W}$, and let $\mathbf{W}^*$ denote the best possible weight matrix for the ADL model

$$\mathbf{W}^* = \arg\min_{\mathbf{W}} \text{MSFE}(\mathbf{W}).$$

Furthermore, we let MSFE(1) denote the MSFE achieved by the MLE. Under correct model specification, we naturally have that MSFE(1)=MSFE($\mathbf{W}^*$). Proposition 4 states a GW test with a null hypothesis of correct specification

$$\text{H}_0 : \text{MSFE}(\mathbf{W}^*) = \text{MSFE}(1) \quad \text{(both MLE and FRW provide equal forecasting accuracy)}$$

against an alternative of incorrect specification

$$\text{H}_1 : |\text{MSFE}(\mathbf{W}^*) - \text{MSFE}(1)| \geq \delta > 0 \quad \text{(FRW provides improved forecast accuracy).}$$

Below we let $\bar{d}_{m,K}(\mathbf{w}_H)$ and σ_K^2 denote the sample average and variance of the n-step-ahead MSFE difference between FRW and MLE, evaluated over K out-of-sample observations, i.e. $\bar{d}_{m,K}(\mathbf{w}_H) = \frac{1}{K}\sum_{k=T-n}^{T^\star} d_{m,k+n}(\mathbf{w}_{H,k})$, with T and $T^\star$ denoting the beginning and the end of the forecast evaluation sample, respectively, and

$$d_{m,k+n}(\mathbf{w}_{k,H}) := u_{m,k+n}(1)^2 - u_{m,k+n}(\mathbf{w}_{H,k})^2,$$

where $u_{m,k+n}(\mathbf{w}_{H,k}) := \hat{Y}_{k+n}\big(\hat{\boldsymbol{\theta}}_{m,k}(\mathbf{w}_{H,k})\big) - Y_{k+n}$.

Similarly to Giacomini and White (2006), we use a suitable HAC-type estimator of the asymptotic variance σ_K^2, for example, $\hat{\sigma}_K^2 = \frac{1}{K}\sum_{k=T-n}^{T^\star} d_{m,k+n}^2 + 2[K^{-1}\sum_{j=1}^{p_K}\omega_{K,j}\sum_{k=T-n+j}^{T^\star} d_{m,k+n}d_{m,k+n-j}]$, with $\{p_K\}$ such that $p_K \to \infty$ as $K \to \infty$, $p_K = o(K)$ and $\{\omega_{K,j} : K = 1,2,\ldots; j = 1,\ldots,p_K\}$ a triangular array such that $|\omega_{K,j}| < \infty$, $K = 1,2,\ldots,p_K$, and $\omega_{K,j} \to 1$ as $K \to \infty$ for each $j = 1,\ldots,p_K$.

PROPOSITION 4. *Let the conditions of Proposition 1 hold and suppose* $0 < \sigma_K^2 < \infty$ *for all* K *sufficiently large. Then*

$$\sqrt{K}\bar{d}_{m,K}(\mathbf{w}_H) \,/\, \hat{\sigma}_K \xrightarrow{d} \mathcal{N}(0,1), \qquad as \qquad K \to \infty,$$

for any given forecast horizon n *and estimation window size* $m < \bar{m} < \infty$ *under the null hypothesis* $H_0 : MSFE(\mathbf{W}^*) = MSFE(1)$*, i.e.* $\mathbb{E}[d_{m,k+n}(\mathbf{w}_{H,k})] = 0$*, and*

$$|\sqrt{K}\bar{d}_{m,K}(\mathbf{w}_H) \,/\, \hat{\sigma}_K| \to \infty, \qquad as \qquad T \to \infty,$$

with probability approaching to one under the alternative hypothesis $\big|\mathbb{E}[d_{m,k+n}(\mathbf{w}_{H,k})]\big| \geq \delta > 0$ *for all* K *sufficiently large.*

5 Monte Carlo evidence

We verify in a set of simulation experiments whether the small-sample properties match our theoretical findings. We investigate the finite sample performance of FRW in the context of the basic AR(1) model, which is a special case of the ADL(p, q) model with $q = 0$, and by means of our Monte Carlo simulation experiments. We consider four data generating processes (DGP) for the time series y_t, these are four departures of the AR(1) model: (1) the basic AR(1) model, (2) the AR(1) model with a time-varying coefficient, and (3) the AR(1) model with a structural break for the AR coefficient. For each case, we assess whether the optimal weight function does improve the one-step-ahead forecasting accuracy of the MLE. The forecasts are based on rolling window estimates. The FRW method is only applied to the standard AR(1) model. We adopt the exponential weight function with $w_{k,t} = \rho^{k-t}$ with $\rho \in [0, 1]$ and $k = T', \ldots, T$. More details for the Monte Carlo experiments and their results are presented in Appendix B. The overall conclusions are as follows. In case the AR model is well-specified as in DGP (1), we do not find any significant improvements in forecasting accuracy. In effect, FRW and MLE results are overall the same. However, in all three other cases where the AR model is not correctly specified for the DGP, the FRW succeeds in producing higher maximized likelihood function values using non-uniform weight functions. In particular, the increases in forecast precisions of FRW compared to MLE are high and they confirm the asymptotic findings in the previous section.

6 Empirical illustrations

We study the empirical performance of our FRW procedure for several economic applications. In Section 6.1, we focus on forecasting the growth rate of monthly U.S. Industrial Production Index (IPI) during the global recession of 2008. We find that the FRW can deliver a significantly better forecasting performance than the standard MLE

method and the optimal rolling window (ORW) method. This result is achieved by increasing the weights associated with observations coming from past recession periods relative to expansion periods. Moreover, we find that the improved out-of-sample forecasting accuracy delivered by FRW is not driven by a single, or just a few, observations. Instead, the improved forecasting performance is experienced throughout the entire evaluation sample. In Section 6.2 we show that the improved forecasting performance delivered by FRW is not restricted to forecasting the IPI during a recession period. In particular, Section 6.2 shows that the FRW outperforms the MLE also during expansion periods. In Section 6.3, we show that the FRW outperforms the MLE in other data sets as well. In effect, it reveals that the FRW method can, in some cases, improve the forecasting performance of the model in a remarkable fashion. Finally, in Section 6.4, we focus on forecasting during the COVID-19 recession. We find that even when simple exponential weights are used the FRW method can improve substantially the forecast accuracy especially in the presence of a structural break.

6.1 Forecasting industrial production during the great recession

In this section, we study whether the FRW method improves the forecast accuracy for the growth rates in the US Industrial Production Index (IPI), when compared to the classical MLE and single rolling-window with optimal window size (ORW) methods. Here, the single rolling window corresponds to the FRW with simple binary weights, as explained in Section 2.1.

The IPI growth rate is a core indicator of the US economy. To evaluate the forecasting performance of the FRW and ORW methods during the great recession, we consider the monthly time series of the US IPI growth rates from January 1950 until June 2009. We consider two window lengths for the estimation of the parameters: 25 years (period 1975–1999, containing 33 recession months) and 50 years (period 1950–1999, containing 85

recession months). For the cross-validation of forecasting accuracy, we have 8 years (period 2000–2007, with 8 recession months). Finally, we take 1.5 years (period 2008–2009 June, containing 18 recession months) for the out-of-sample forecasting period. We consider simple AR models as well as ADL models where the latter includes exogenous regressors. For the ADL model, the set of exogenous predictors consists of the growth rates of the IP sectoral-level series: manufacturing, mining, and utilities. We report the results for the forecast horizons of one month ($n = 1$) and one quarter ($n = 3$) ahead, focusing on direct forecasts.

Since IPI observations coming from past recession periods may be more informative about the dynamics of the IPI during the global recession of 2008–2009, in this empirical application, we use a weighting scheme that can potentially give higher weight to the recession periods as described in Section 3.3. Specifically, the weight function for $\mathbf{w}_T$ takes the form

$$w_t = \rho_2 \cdot (1 - Z_t) + (1 - \rho_2) \cdot Z_t, \quad 0 \leq \rho_2 \leq 1,$$

where the indicator variable $Z_t \in \{0, 1\}$ is the NBER recession indicator. Naturally, when $\rho_2 < 0.5$, observations from recession periods receive more weight in estimation. The ratio $(1 - \rho_2)/\rho_2$ indicates how much larger the weights for recession periods are compared to expansion periods. The coefficient ρ_2 is estimated via cross-validation using Algorithm 1, allowing the data to determine the strength of this effect. Since the NBER recession indicator is released with a significant delay, it can be replaced by the recession forecast $\hat{Z}_t$ when unavailable. This alternative approach is explored in Supplementary Appendix C.

The weight function can be made even more flexible by allowing for a combination of a binary weight, ρ_2, and an exponentially decaying weight, ρ_1, as in (5). This weighting scheme puts emphasis on the recession periods as well as downweighs observations from the remote past. The intuition that underlies the exponential decay is straightforward: as the

economy changes due to institutional, technological and societal shifts, observations far in the past are likely to be less informative about the current dynamics of IPI than the more recent observations. Setting $\rho_1 = 1$ and $\rho_2 = 0.5$ delivers classical MLE.

For simplicity, we set $\rho_1 = 1$ and apply the FRW method with an NBER recession weight function to the AR(1) and AR(p) models, as well as the ADL$(1,1)$ and ADL(p,q) models, where the number of lags p and q are selected by means of the Akaike's Information Criterion (AIC). We refer to the latter case as the AR(AIC) and ADL(AIC) models. We focus on out-of-sample forecasting of the monthly US IPI growth rate during the NBER global recession period, which spans period from January 2008 to June 2009. We emphasize that the sample for which the forecasting accuracy was evaluated was not used for estimating the AR/ADL model parameters or the likelihood weights. Hence, the MLE estimator can potentially outperform the FRW estimator in this separate evaluation sample. Any improvement in forecasting accuracy should only be expected if the FRW method presents a substantial estimation advantage over the MLE.

The results are presented in Table 1. We take the recession period in the year 2001 as our cross-validation period to determine the feature of the FRW (ρ_2) and optimal rolling window size of the single rolling window. The weights are further used to produce forecasts during the global recession of 2008. We report the RMSE ratios for the FRW against the MLE and ORW methods such that a ratio value smaller than unity indicates improvements in forecast accuracy. We also report the p-values of the GW test (in parentheses) and the cross-validation estimates of the FRW parameter ρ_2 (multiplied by 100) and the optimal window size. The optimal window size was determined using a grid search with a minimum window size of 4 years (48 months) and a step size of 1 quarter (3 months).

The results from Table 1 reveal that the FRW strongly outperforms both the MLE and ORW in the majority of cases. The improvements of up to 34% in the forecasting accuracy of the AR(AIC) relative to MLE are quite remarkable. The p-values of the GW

Table 1: Pseudo out-of-sample forecasts: Binary NBER weights for recessions.

		RMSE				MAE			
		n=1		n=3		n=1		n=3	
		Ratio (GW)	ρ_2	Ratio (GW)	ρ_2	Ratio (GW)	ρ_2	Ratio (GW)	ρ_2
AR(1)									
FRW/MLE	25 years	1.116	2.8	0.842***	3.1	0.866	1.3	0.719***	1.8
		(0.292)		(0.009)		(0.216)		(0.001)	
	50 years	0.917**	6.1	0.806***	8.3	0.785***	3.0	0.656***	6.9
		(0.017)		(0.001)		(0.000)		(0.000)	
FRW/ORW	50 years	0.880**	333	0.856***	327	0.738***	330	0.739***	327
		(0.026)		(0.001)		(0.000)		(0.000)	
AR(AIC)									
FRW/MLE	25 years	1.219	3.0	0.870	3.1	1.167	0.6	0.820	1.8
		(0.169)		(0.131)		(0.266)		(0.160)	
	50 years	0.938**	8.0	0.806***	8.3	0.836***	3.9	0.656***	6.9
		(0.045)		(0.001)		(0.006)		(0.000)	
FRW/ORW	50 years	0.800***	54	0.856***	327	0.705***	51	0.739***	327
		(0.006)		(0.001)		(0.001)		(0.000)	
ADL(1,1)									
FRW/MLE	25 years	1.097	3.5	1.316	4.8	1.014	3.3	1.186	3.2
		(0.277)		(0.139)		(0.477)		(0.224)	
	50 years	0.854***	5.5	0.818***	7.8	0.734***	4.0	0.704***	6.1
		(0.008)		(0.006)		(0.000)		(0.000)	
FRW/ORW	50 years	0.867**	330	0.858**	327	0.778***	330	0.767***	327
		(0.020)		(0.019)		(0.005)		(0.001)	
ADL(AIC)									
FRW/MLE	25 years	1.095	5.5	0.972	4.2	1.208	3.3	1.007	2.2
		(0.173)		(0.384)		(0.169)		(0.482)	
	50 years	0.870**	8.1	0.805**	9.1	0.771***	4.4	0.697***	6.4
		(0.027)		(0.011)		(0.001)		(0.000)	
FRW/ORW	50 years	0.995	54	0.811*	51	0.936	54	0.601***	48
		(0.481)		(0.080)		(0.375)		(0.008)	

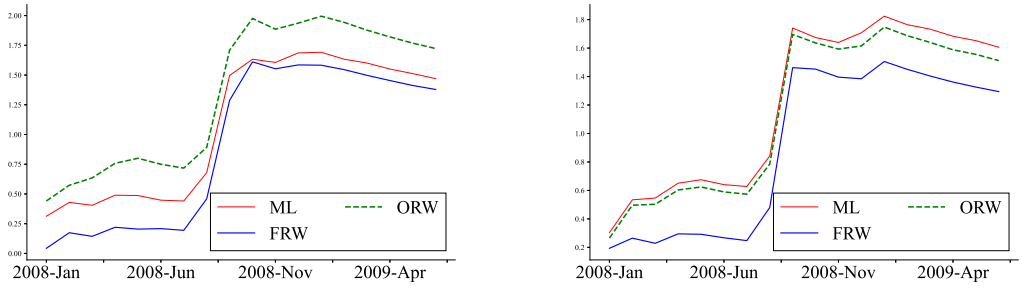
Table 1: *FRW* forecasting results for monthly U.S. IPI growth rate for the AR(1) and AR(p) models, as well as ADL$(1,1)$ and ADL(p,q) models, where p and q are selected by the Akaike's information criterion, AR(AIC) and ADL(AIC). Estimation window is of size 25 and 50 years, starting from periods 1975-1999 and 1950-1999, respectively. The column labeled Ratio shows the RMSE using the FRW method of the AR/ADL models relative to the RMSE using the MLE or ORW methods. Cases where the forecasting improvement is statistically significant at the 90%, 95%, and 99% confidence levels are indicated by *, **, and ***, respectively. Entries in parentheses show the p-value of GW test. The column labelled "ρ_2" contains either estimated optimal window size or the optimal features of the FRW, multiplied by 100, for the binary weighting $\rho_2 \cdot (1 - Z_t) + (1 - \rho_2) \cdot Z_t$ where $Z_t \in \{0, 1\}$ is the NBER recession indicator. The estimation of the FRW features is performed on the cross-validation sample once. The cross-validation sample is recession period of 2000–2007. The out-of-sample forecasts are computed for 2008–2009 June (18 months). The optimal window size was determined based on a grid search with a minimum window size of 4 years (48 months) and a step size of 1 quarter (3 months).

test statistics reveal that forecast accuracy improvements relative to MLE and ORW are statistically significant at standard confidence levels. Most importantly, the parameter estimates of ρ_2 indicate that the improved performance is due to giving more weight to past recession periods. We want to highlight that the improved forecasting performance is not only achieved for the simple AR(1) model, which is likely to be a misspecified model, but also for the AR(AIC) and ADL(AIC) models, which in many cases has $p > 1$ (and $q > 1$) in the AR(p) (ADL(p, q)) model and is therefore more flexible as it includes more lags. Moreover, the FRW outperforms the ORW where the latter uses only a part of the sample for estimation, which should make the parameter estimates more prone, for example, to the changes in the dynamics of the time series and consequently to superior forecasts. Hence, the improvements of the FRW over the ORW highlight the importance of allowing for more flexible weighting scheme. Finally, the FRW demonstrates superior forecasting performance over both the MLE and the ORW in the cases of the ADL$(1, 1)$ and ADL(AIC) models.

The cross-validation estimates of ρ_2 vary over a range of values smaller than 0.09. The more reasonable, large estimates of ρ_2 lead to better results than the more extreme ones. For example, consider the AR(AIC) model and the rolling window estimation sample of 50 years, for the RMSE of the quarterly forecasting horizon ($n = 3$) we have the estimate of 0.083 for ρ_2. The estimate of 0.083 implies that observations coming from a recession period receive 11 times more weight than observations coming from an expansion period. Hence, an observation in the recession year of 2001 is given approximately the same weight as an observation in the 11 expansion years before 2000. Most importantly, this weighting scheme leads to significant improvements in the RMSE as highlighted by the ratio of 0.806. In contrast, the ρ_2 estimate of 0.006 (obtained for the MAE and for the monthly forecast horizon ($n = 1$) and the rolling window of 25 years) implies essentially that only the observations in recession periods have an impact on the estimation; the data coming from

expansion periods are ignored. Such extreme estimates of ρ_2 lead to poor results in out-of-sample forecasting evaluations when compared to the standard MLE method.

To illustrate that the results reported in Table 1 are not driven by only a few observations, we present in Figure 3 the accumulated RMSE statistics over the 18 months of our forecasting window of 2008 January to 2009 June, for the MLE, ORW, and FRW methods. Figure 3 shows these results for AR(AIC) model for both $n = 1$ and $n = 3$. They clearly show that the performance of the FRW method is preferred, in terms of RMSE, for all time periods in the forecasting window. It is a strong empirical result that the FRW achieves better forecasting accuracy by outperforming the MLE and ORW over the entire forecasting evaluation sample.



$n = 1$ $n = 3$

Figure 3: The accumulated root mean squared forecasting error from the MLE, ORW, and FRW methods during the forecasting window of 2008–2009 June. These results are presented for AR(AIC) model with a 50-year rolling window. The forecasts are computed for a forecasting horizons of $n = 1$ and $n = 3$ periods.

6.2 Forecasting industrial production during the 2008-expansion

Given the encouraging results as reported above, we have concluded that the FRW method can lead to improved forecasting performances for IPI growth during recession periods. We next investigate whether these improvements are not specific to the global financial recession period only. Does the FRW method also provide significant improvements in forecasting accuracy during expansion periods? For this purpose, we take the NBER expansion period

Table 2: AR model forecasting: for Expansion period,
with exponential decay (ρ_1) and binary (ρ_2) weights

	n=1			n=3		
	Ratio (GW)	ρ_1	ρ_2	Ratio (GW)	ρ_1	ρ_2
AR(1)	0.982 (0.328)	0.934	100.0	0.979 (0.112)	1.000	100.0
AR(AIC)	0.969* (0.061)	1.000	100.0	0.976* (0.083)	1.000	100.0

Table 2: FRW forecasting results for monthly U.S. IPI growth rate. The ratios are based on the RMSE statistics and a rolling window size of 25 years. The forecasts are computed for the expansion period 2002–2007. For further explanations, we refer to Table 1.

from January 2002 to December 2007 as our forecasting evaluation sample. The results for AR(1) and AR(AIC) models are reported in Table 2 that only focuses on the RMSE ratios and a rolling window size of 25 years; the full results, including the MAE ratios and the rolling window size of 50 years, are provided in Appendix C. We learn from Table 2 that the FRW method delivers higher out-of-sample forecasting precisions in the evaluation sample when compared to those from the MLE method. In two of these cases, the GW p-values show improvements that are marginally statistically significant at the 90% confidence level.

In particular, Table 2 presents overall improvements in forecasting accuracy of up to 3%. Although the magnitude of these improvements is not as impressive as in the recession case, we still find statistically significant improvements in terms of the GW test. The improvements are especially relevant when related to the forecasts of the AR(AIC) model, which is a highly competitive benchmark in many forecasting studies. In any case, improvements in forecasting accuracy are expected to be smaller in the expansion case rather than in the recession case. The main reason for this is that the number of observations originating from economic expansion periods outnumbers by far the number of observations originating from recession periods. As a result, the total weight that expansion

periods have in the likelihood criterion is much larger. It leads to ML parameter estimates that are already adequate for expansion periods. As a result, there is a smaller margin for obtaining an improvement for the FRW method. This also explains the finding that the weights that only weigh expansions periods are selected ($\rho_2 = 1$), effectively disregarding all observations from recession periods. In Appendix C, we demonstrate that the conclusion remains unchanged when the expansion period in the aftermath of the global recession is considered.

6.3 Forecasting other variables during the 2008-recession

To show that the FRW method can also be effective for other relevant economic time series in delivering notable improvements in forecasting accuracy, we present two additional illustrations. We consider the monthly economic time series of the U.S. Unemployment Rate and U.S. Total Non-Farm Payrolls growth rates and focus again on forecasting these variables during the global recession of 2008. In Table 3 we report the RMSE and MAE ratios obtained from the same out-of-sample forecasting study as done for U.S. Industrial Production. The tables with full estimation results for the two time series, including the estimates for ρ_2, are presented in Appendix C.

In the upper part of Table 3, the RMSE and MAE ratios are presented for the monthly time series of the US Unemployment Rate. These results show that the FRW is capable of delivering significant reductions in the out-of-sample forecasts. In particular, we report RMSE reductions of up to 39% and MAE reductions of up to 36% for the forecasts from the AR(1) model, both at a monthly horizon. The lower part of Table 3 is for the forecasting of the monthly time series of the U.S. Non-Farm Payrolls growth rates during the global recession. For the AR(1) model, remarkable reductions of up to 38% in the out-of-sample forecasting RMSE are obtained, and reductions of more than 50% in the MAE. Even more convincingly, for the AR(AIC) we find reductions of 25% in the RMSE and around 42% in

the MAE. Most of these reductions are statistically significant at any reasonable confidence level.

Table 3: AR model forecasting: Other economic time series

		RMSE Ratio		MAE Ratio	
		n=1	n=3	n=1	n=3
		U.S. Monthly Unemployment Rate			
FRW/MLE	AR(1)	0.608**	0.617***	0.639**	0.692***
		(0.019)	(0.007)	(0.028)	(0.001)
	AR(AIC)	0.637**	0.691**	0.674*	0.706***
		(0.030)	(0.010)	(0.056)	(0.006)
FRW/ORW	AR(1)	0.671***	0.700***	0.781**	0.845***
		(0.005)	(0.003)	(0.011)	(0.000)
	AR(AIC)	0.740**	0.746***	0.765*	0.844***
		(0.021)	(0.009)	(0.050)	(0.004)
		U.S. Monthly Total Non-farm Payrolls			
FRW/MLE	AR(1)	0.623***	0.734*	0.423***	0.622**
		(0.000)	(0.078)	(0.000)	(0.028)
	AR(AIC)	0.755**	0.855	0.584***	0.715*
		(0.033)	(0.236)	(0.004)	(0.091)
FRW/ORW	AR(1)	0.956	0.795***	0.887	0.728***
		(0.368)	(0.003)	(0.256)	(0.004)
	AR(AIC)	0.840**	0.593***	0.794**	0.526***
		(0.022)	(0.007)	(0.037)	(0.001)

Table 3: FRW forecasting results for two other economic time series. The forecasts are computed over the sample period 2008 January - 2009 June. We refer to Table 1 for further explanations.

6.4 Forecasting during the COVID-19 recession

We also study the forecasting performance of the FRW method during the COVID-19 recession. The COVID-19 global recession is one of the deepest recessions in the world history and it is characterized by a huge decline in economic activity in many sectors as well as increased uncertainty. Therefore, in our application, we consider two time series: the U.S. Economic Policy Uncertainty Index (EPUI) proposed by Baker et al. (2016) and the changes over the month in the U.S. Weekly Economic Index (WEI) constructed by Lewis et al. (2022). The former is based on the U.S. newspapers and it characterizes policy uncertainty, while the latter is constructed using ten indicators of the U.S. real economic activity which cover consumer behavior, labor market, and production.

For the daily time series of the U.S. EPUI, we take the estimation and cross-validation samples spanning from 1 February 2018 until 1 February 2020, and for the U.S. Weekly Economic Index (WEI) from 5 January 2008 until 1 February 2020 with the rolling windows' length of 550 days and 533 weeks, respectively. Then for the cross-validation, we have 180 days (around half a year) for the EPUI and 96 weeks (around 2 years) for the WEI. Starting from 1 February 2020, right before the recession began, we then use 121 days (3 months) and 33 weeks, respectively, for the out-of-sample evaluation. To compare the forecasting performance we consider one-step-ahead forecasts ($n = 1$) and exponentially decaying weights for the FRW.

In Table 4 we report the RMSE and MAE ratios for the FRW against the MLE method. We find that for the U.S. EPUI series the FRW method substantially outperforms the standard MLE method and the improvements are of up to 24%. Moreover, this result is significant at a 1% level. These improvements most likely occur due to the presence of a structural break in the series as during the coronavirus pandemic the policy uncertainty increased substantially since the middle of March 2020, shortly after the recession started.

Therefore, the data far in the past became less relevant for forecasting and receive much lower weights since the end of March 2020. Furthermore, in Figure 4, the accumulated RMSEs for MLE and FRW are similar in February but, after the break, the accumulated RMSE based on the MLE is much higher, which again supports the idea that after the break the AR model is misspecified and then the FRW method is preferable for forecasting. This is in line with the findings of the Monte Carlo experiment B.3 with a structural break. The performance of the ORW and FRW is comparable as indicated by the ratio in Table 4 and accumulated RMSEs in Figure 4. Intuitively, when optimal window size is low, the ORW also limits the contribution of the observations far in the past. Hence, in the presence of a structural break, both FRW and ORW deliver good forecasting results.

For the U.S. WEI, the FRW method slightly outperforms the standard MLE based on the RMSE but the difference in the forecast accuracies is not significant (Table 4). In Figure 4 we also observe that at one moment the FRW outperforms MLE but overall their performance is comparable. This result could be explained by the fact that the series exhibited a large downturn at the end of March and after that quickly started bouncing back. The FRW method reacts quicker to the downturn but then the difference in forecast accuracies disappears. This is in line with the theoretical findings discussed in Section 4 that in some cases the FRW method simplifies to a standard MLE. Overall, we can conclude that the use of the FRW method appears to be highly beneficial for accurate economic forecasting.

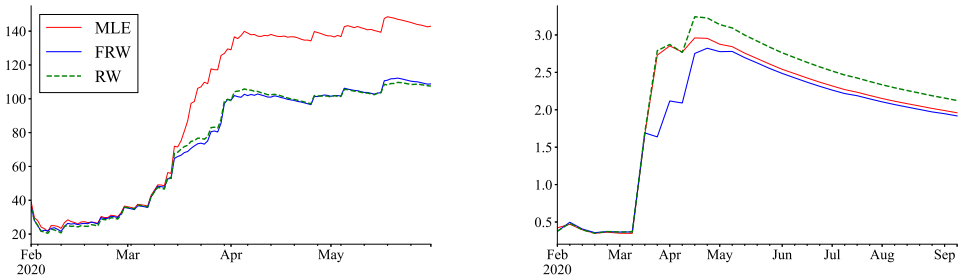
7 Conclusion

We have introduced a new flexible rolling-window estimation (FRW) method that weighs the likelihood function contributions of individual observations differently for the purpose to deliver optimal forecasting accuracy for linear autoregressive models and autoregressive

Table 4: AR model forecasting: COVID-19 recession

	RMSE Ratio n=1	Avg(ρ)	MAE Ratio n=1	Avg(ρ)
	U.S. Daily Economic Policy Uncertainty Index			
FRW/MLE	0.761***	0.854	0.794***	0.842
	(0.000)		(0.000)	
FRW/ORW	0.993		1.006	
	(0.785)		(0.487)	
	Δ_4 *U.S. Weekly Economic Index*			
FRW/MLE	0.979	0.880	0.992	0.888
	(0.447)		(0.459)	
FRW/ORW	0.903**		0.967**	
	(0.048)		(0.016)	

Table 4: FRW forecasting results for two economic time series during the COVID-19 recession. The forecasts are computed over the sample period 1 February 2020–31 May 2020 and 1 February 2020–12 September 2020 for the EPUI and WEI, respectively. We refer to Table 1 for further explanations.



EPUI WEI

Figure 4: The accumulated root mean squared forecasting error for the MLE, ORW, and FRW methods during the forecasting window of 1 February 2020–31 May 2020 for the EPUI and 1 February 2020–12 September 2020 for the WEI. The forecasts are computed for a forecasting horizon of $n = 1$ periods.

distributed lag models. We have shown how to estimate the optimal weights using a cross-validation method. We further have investigated the asymptotic properties of FRW and we have considered four Monte Carlo experiments for studying the finite-sample properties. In empirical illustrations, we have analyzed the forecast accuracy of the FRW method compared to standard MLE for several key economic indicators during the global recession

of 2008–2009, the 2008 expansion, and the COVID-19 recession. The analyses have revealed that the FRW can substantially improve forecast accuracy. In particular, forecast precision during recession periods can be significantly improved by increasing the weights of the most recent observations or by increasing the weights of observations corresponding to similar recession periods in the past. We have made the case that econometricians may need to give past recessions more attention for providing more accurate forecasts during periods of recessions and financial crises. In further work we can extend the method towards other linear dynamic models, including the autoregressive moving average model, and towards multivariate specifications such as the vector autoregressive model. Nonlinear extensions of the model classes can also be considered but may require an extended theoretical foundation.

References

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics 131*(4), 1593–1636.

Blasques, F., S. J. Koopman, and A. Lucas (2015). Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika 102*(2), 325 – 343.

Boud, K., A. De Block, G. Langenus, and P. Reusens (2023). Nowcasting GDP through the lens of economic states. *National Bank of Belgium 445*.

Clements, M. and D. F. Hendry (1998). *Forecasting economic time series*. Cambridge University Press.

Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians*. Oxford University Press.

Dendramis, Y., G. Kapetanios, and M. Marcellino (2020). A similarity-based approach for macroeconomic forecasting. *Journal of the Royal Statistical Society Series A: Statistics in Society 183*(3), 801–827.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–63.

Feng, S. and J. Sun (2020). Misclassification-errors-adjusted Sahm rule for early identification of economic recession. *IZA Discussion Paper 13168*.

Giacomini, R. and B. Rossi (2009). Detecting and predicting forecast breakdowns. *Review of Economic Studies 76*(2), 669–705.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Giraitis, L., G. Kapetanios, and S. Price (2013). Adaptive forecasting in the presence of recent and ongoing structural change. *Journal of Econometrics 177*(2), 153–170.

Goyal, A. and I. Welch (2003). Predicting the equity premium with dividend ratios. *Management Science 49*(5), 639–654.

Inoue, A., L. Jin, and B. Rossi (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics 196*(1), 55–67.

Koop, G. and S. Potter (2004). Forecasting in dynamic factor models using Bayesian model averaging. *Econometrics Journal 7*(2), 550–565.

Lewis, D., K. Mertens, J. Stock, and M. Trivedi (2022). Measuring real activity using a weekly economic index. *Journal of Applied Econometrics 37*(4), 665–840.

Molodtsova, T. and D. H. Papell (2009). Out-of-sample exchange rate predictability with Taylor rule fundamentals. *Journal of International Economics 77*(2), 167–180.

Oh, D. H. and A. J. Patton (2024). Better the devil you know: Improved forecasts from imperfect models. *Journal of Econometrics 242*(1), 105767.

Paye, B. S. and A. Timmermann (2006). Instability of return prediction models. *Journal of Empirical Finance 13*(3), 274–315.

Pötscher, B. M. and I. R. Prucha (1997). *Dynamic nonlinear econometric models: Asymptotic theory*. Springer.

Rossi, B. (2013). Advances in forecasting under instability. In *Handbook of Economic Forecasting*, Volume 2, pp. 1203–1324. Elsevier.

Sahm, C. (2019). Direct stimulus payments to individuals. *Recession ready: Fiscal policies to stabilize the American economy*, 67–92.

Schinasi, G. J. and P. A. V. B. Swamy (1989). The out-of-sample forecasting performance of exchange rate models when coefficients are allowed to change. *Journal of International Money and Finance 8*(3), 375–390.

Stock, J. H. and M. W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics 14*(1), 11–30.

Stock, J. H. and M. W. Watson (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking 39*(s1), 3–33.

Swanson, N. R. (1998, May). Money and output viewed through a rolling window. *Journal of Monetary Economics 41*(3), 455–474.

Tarassow, A. (2019). Forecasting U.S. money growth using economic uncertainty measures and regularisation techniques. *International Journal of Forecasting 35*(2), 443–457.

Wolff, C. C. P. (1987). Time-varying parameters and the out-of-sample forecasting performance of structural exchange rate models. *Journal of Business & Economic Statistics 5*(1), 87–97.

Technical Appendix

A Proofs of propositions

Proof of Proposition 1. Theorem 6.10 in Pötscher and Prucha (1997) implies that the sequence $\{Y_t\}_{t\in\mathbb{N}}$ initialized at $Y_1 = y$ and generated according to (8) for every $t \in \mathbb{N}$ is NED of size $-a$ as long as $\{V_t\}_{t\in\mathbb{Z}}$ is a NED n_V-variate stochastic sequence of size $-a$, $\sup_t \mathbb{E}|V_t|^2 < \infty$, $\sup_{v\in\mathcal{V}} \sup_{y\in\mathcal{Y}} |\phi'_v(y,v)| < \infty$ and $\sup_{v\in\mathcal{V}} \sup_{y\in\mathcal{Y}} |\phi'_y(y,v)| < 0$. Given conditions *(i)-(v)*, all the conditions of the theorem are satisfied. The bounded moments of $\{Y_t\}_{t\in\mathbb{N}}$ are also ensured by condition (iii) in Pötscher and Prucha (1997, Theorem 6.10). The result for the limit sequence $\{Y_t\}_{t\in\mathbb{Z}}$ then follows straightforwardly.

The ADL(p,q) representation follows trivially by rewriting the Y_t as follows

$$\begin{aligned} Y_t &= \phi(Y_{t-1}, V_t) - \epsilon_t + \epsilon_t \\ &= \frac{\phi(Y_{t-1}, V_t) - \epsilon_t}{\psi_0 + \Psi(L)Y_t + K(L)X_t}(\psi_0 + \Psi(L)Y_t + K(L)X_t) + \epsilon_t \,, \end{aligned}$$

where $\Psi(L)$ and $K(L)$ denote the lag polynomials $\Psi(L) = \psi_1 L + \cdots + \psi_p L^p$ and $K(L) = \kappa_1 L + \cdots + \kappa_q L^q$, and finally defining

$$\begin{aligned} \alpha_{i,t} &:= \frac{\big(\phi(Y_{t-1}, V_t) - \epsilon_t\big)}{\psi_0 + \Psi(L)Y_t + K(L)X_t}\psi_i \,, \qquad \text{for } i = 0, 1, ..., p, \\ \beta_{j,t} &:= \frac{\big(\phi(Y_{t-1}, V_t) - \epsilon_t\big)}{\psi_0 + \Psi(L)Y_t + K(L)X_t}\kappa_j \,, \qquad \text{for } j = 1, ..., q. \end{aligned}$$

The NED nature of $\alpha_{i,t}$, $i = 0, ..., p$ and $\beta_{j,t}$, $j = 1, ..., q$, follows from Davidson (1994, Theorem 17.12) since every $\alpha_{i,t}$ and $\beta_{j,t}$ is a Lipschitz function of NED variables. □

Proof of Proposition 2. Since the Gaussian ADL(p,q) model is well specified, it follows immediately that the Gaussian MLE converges to the true parameter as $T' \to \infty$,

i.e. $\hat{\boldsymbol{\theta}}_T(1) \xrightarrow{p} \boldsymbol{\theta}_0^*(1) = \boldsymbol{\theta}_0$ under the usual regularity conditions.

Application of a continuous mapping theorem as $T' \to \infty$ implies that

$$u_t(\hat{\boldsymbol{\theta}}_T(1))^2 \equiv (\hat{Y}_{k+n}\big(\hat{\boldsymbol{\theta}}_T(1)\big) - Y_{k+n})^2 \xrightarrow{p} u_t(\boldsymbol{\theta}_0^*(1))^2 \equiv (\hat{Y}_{k+n}\big(\boldsymbol{\theta}_0\big) - Y_{k+n})^2.$$

As $T' \to \infty$, the limit of the n-step-ahead MSFE forecast criterion based on H observed forecast errors under the true parameter $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_0^*(1)$ is given by

$$Q_H(\mathbf{1}) := \frac{1}{H}\sum_{t=1}^{H} u_t(\boldsymbol{\theta}_0^*(1))^2.$$

Since $u_t(\boldsymbol{\theta}_0)$ is a Lipschitz function of the NED sequence $\{\epsilon_t\}_{t\in\mathbb{Z}}$, by Theorem 17.12 (Davidson (1994)) the sequence $\{u_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is also NED, and an application of the LLN for NED sequences (Pötscher and Prucha, 1997, Theorem 6.4) then yields

$$Q_H(\mathbf{1}) \xrightarrow{p} \mathbb{E}u_t(\boldsymbol{\theta}_0^*(1))^2 \quad \text{as} \quad H \to \infty.$$

Finally, note that Algorithm 1 is always initialized at a weight matrix $\mathbf{W}$ satisfying $\hat{\mathbf{w}}_{k,t} = 1 \;\forall\; k$ and $\forall\; t \leq k$. As a result, in the limit as $T' \to \infty$ and $H \to \infty$, the probability that $\hat{\mathbf{w}}_{k,t} \neq 1$ for some (k,t) is given by

$$\mathbb{P}(\hat{\mathbf{w}}_{k,t} \neq 1) = \mathbb{P}\Big(\mathbb{E}u_t(\boldsymbol{\theta}_0^*(\mathbf{w}_k))^2 < \mathbb{E}u_t(\boldsymbol{\theta}_0^*(1))^2\Big) = 0$$

for any weight vector $\mathbf{w}_k$ with some element $w_{k,t} \neq 1$ and some pair (t,k). □

Proof of Proposition 3. The first claim follows trivially from the design of Algorithm 1. The second claim follows by noting that both the ADL parameter vector $\hat{\boldsymbol{\theta}}$ and the weight

parameter vector $\hat{\boldsymbol{\rho}}$ satisfy a recursive relation

$$\hat{\boldsymbol{\theta}}_T^{(j+1)} = \hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\rho}}_H(\hat{\boldsymbol{\theta}}_T^{(j)})) \quad \text{and} \quad \hat{\boldsymbol{\rho}}_H^{(j+1)} = \hat{\boldsymbol{\rho}}_H(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\rho}}_H^{(j)})) \quad \forall\ j \geq 1.$$

It is well known that the uniform unit bound on the derivative ensures the stability of the recursion towards a unique global fixed point for any initialization. □

Proof of Proposition 4. First, we show that under the null hypothesis $\sqrt{K}\bar{d}_{m,K}/\sigma_K \xrightarrow{d} \mathcal{N}(0,1)$, where $\sigma_K^2 = \mathbb{V}ar(\sqrt{K}\bar{d}_{m,K})$ and $0 < \sigma_K^2 < \infty$.

We have $\sqrt{K}\bar{d}_{m,K}/\sigma_K = K^{-1/2}\sum_{k=T-n}^{T^\star} \sigma_K^{-1} d_{m,k+n}$. Below, we verify that the sequence $\{\sigma_K^{-1} d_{m,k+n}\}_{k\in\mathbb{Z}}$ satisfies the conditions of the CLT for NED sequences (Pötscher and Prucha, 1997, Theorem 10.2).

Since by Proposition 1 the sequence $\{(Y_t, X_t)\}_{t\in\mathbb{Z}}$ is NED of size $-2(r-1)/(r-2)$, which implies that it is also NED of size -1 as $r > 2$, then by Theorem 17.12 of Davidson (1994) $\{d_{m,k+n}\}_{k\in\mathbb{Z}}$ is also NED of size $-2(r-1)/(r-2)$ since it is Lipschitz on (Y_t, X_t). Moreover, under H_0, we have $\mathbb{E}[d_{m,k+n}] = 0$. Finally,

$$\begin{aligned}
\sup_k \mathbb{E}|\sigma_K^{-1} d_{m,k+n}|^r &= \sigma_K^{-r} \sup_k \mathbb{E}|u_{m,k+n}(1)^2 - u_{m,k+n}(\hat{\mathbf{w}}_{H,k})^2|^r \\
&\leq \sigma_K^{-r}\big(c \sup_k \mathbb{E}|u_{m,k+n}(1)|^{2r} + c \sup_k \mathbb{E}|u_{m,k+n}(\hat{\mathbf{w}}_{H,k})|^{2r}\big) \\
&\leq \sigma_K^{-r} c\big(\sup_k \mathbb{E}|\hat{Y}_{k+n}(\hat{\boldsymbol{\theta}}_{m,k}(1)) - Y_{k+n}|^{2r} \\
&+ \sup_k \mathbb{E}|\hat{Y}_{k+n}(\hat{\boldsymbol{\theta}}_{m,k}(\hat{\mathbf{w}}_{H,k})) - Y_{k+n}|^{2r}\big) < \infty,
\end{aligned}$$

where the last inequality follows since by Proposition 1 $\sup_t \mathbb{E}|Y_t|^{4r} < \infty$ and $\sup_t \mathbb{E}|X_t|^{4r} < \infty$ and ADL forecasts are linear in Y_t and X_t. Hence, the conditions of Theorem 10.2 of Pötscher and Prucha (1997) are satisfied and $\sqrt{K}\bar{d}_{m,K}/\sigma_K \xrightarrow{d} \mathcal{N}(0,1)$.

Now, we show that $\hat{\sigma}_K - \sigma_K \xrightarrow{P} 0$ as $K \to \infty$. As shown above, the sequence $\{d_{m,k+n}\}_{k\in\mathbb{Z}}$

is NED of size $-2(r-1)/(r-2)$ and, under the null, it is also a mean zero process. We also have $\sup_k \mathbb{E}|d_{m,k+n}|^{2r} < \infty$, since by Proposition 1 $\sup_k \mathbb{E}|X_k|^{4r} < \infty$ and $\sup_k \mathbb{E}|Y_k|^{4r} < \infty$. The HAC estimator weights $\omega_{K,j}$ are bounded and $\lim_{K\to\infty} \omega_{K,j} = 1$ for each $j > 0$, then the conditions of Corollary 12.2 of Pötscher and Prucha (1997) are satisfied, and we conclude that $\hat{\sigma}_K - \sigma_K \xrightarrow{P} 0$. Slutsky's theorem finishes the first part of the proof about the behavior of the statistics under the null hypothesis.

For the behavior of the test statistics under the alternative, by the arguments given above, we notice that $\{d_{m,k+n}\}_{k\in\mathbb{Z}}$ is NED of size $-1/2$ and $\sup_k \mathbb{E}|d_{m,k+n}|^{2r} < \infty$. Therefore, by LLN for NED processes (Pötscher and Prucha, 1997, Theorem 6.4), it implies that $\bar{d}_{m,K} - \mathbb{E}[\bar{d}_{m,K}] \xrightarrow{P} 0$. The rest of the proof follows similar reasoning as the proof of Theorem 4 (b) in Giacomini and White (2006).

□

B Simulation experiments

We investigate the finite sample performance of the FRW method in the context of AR(1) model using Monte Carlo simulation. We consider four different data generating processes (DGP) for the time series y_t. Among these four different AR(1) processes, the autoregressive coefficient could be time-invariant, time-varying, regime switching or subject to a structural break. We aim to investigate in which cases the optimal weight function can improve the forecasting accuracy of the MLE. We concentrate on one-step-ahead forecasts based on rolling-window method with window length k. Although we consider four different data generating processes in the simulation experiment, our FRW method is based on the ordinary AR(1) model as given by

$$y_t = \alpha + \beta y_{t-1} + \epsilon_t, \tag{10}$$

where intercept α and autoregressive coefficient β are treated as fixed and unknown, and where ϵ_t is an i.i.d. normally distributed variable with mean zero and unknown variance σ_ϵ^2.

We consider two kinds of weight functions which we apply in Monte Carlo experiment: the exponential weight function ($w_t = \rho_1^{k-t}$, where $\rho \in [0, 1]$) and the binary weight function with decay ($w_t = \rho_1^{(k-t)}(1+(\rho_2-1)\cdot Z_t)$, where Z_t is the predetermined indicator for recession period). The exponential weight function is applied in the first three experiments; while the binary weight function with decay is applied in the last experiment.

B.1 Experiment 1: Time-Invariant AR(1) model

In the first experiment, the data are generated by an AR(1) model with time-invariant parameters and the FRW parameter, ρ_1, is estimated based on Equation (10). The data generating process (DGP) is specified by:

$$\text{DGP: } y_t = \alpha + \beta y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{11}$$

where $\alpha = 0.13$, $\beta = 0.5$ and ϵ_ts are i.i.d distributed with variance $\sigma_\epsilon^2 = 0.5$. The exponential weight function is considered when applying FRW algorithm and we use Equation (10) with the estimates, $\hat{\boldsymbol{\theta}}_T(\mathbf{w}_k(\hat{\boldsymbol{\rho}_H}))$, from the FRW to calculate the forecasts.

In Experiment 1, we generate a time series by the AR(1) model. The rolling window length, k, is selected to be 120 and forecasts are made for time period $t = 701 \ldots 760$. The generated data are considered as monthly data. For each simulated data set, we carry out FRW and use the resulting estimates to calculate the one-step-ahead forecasts. Since the model is accurately specified we expect that the estimated FRW parameter, ρ_1, is close to 1, which means the original AR(1) model can already provide accurate forecasts.

The left panel of Figure 5 shows the simulation density of FRW parameter, ρ_1, in

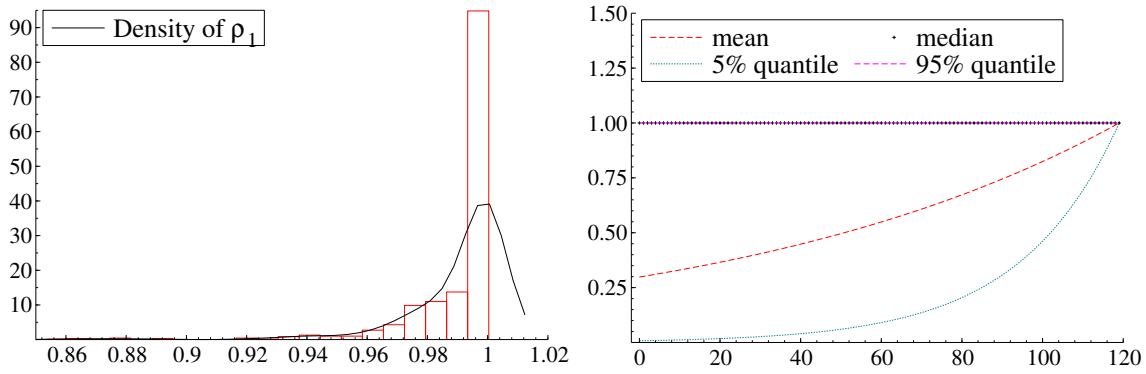


Figure 5: Simulation results for Experiment 1. The data are generated by an AR(1) model and the FRW parameter, ρ_1, is estimated based on Equation (10). Exponential weight function is applied. The left panel presents the simulation density of FRW parameter, ρ_1, over 1,000 simulations. The rights panel presents the average sample weights and its 90% confidence bound.

Experiment 1. The simulated parameters peak at 1. To be precise, the simulated parameter mean is 0.9953 and the median is 1. In the setup, we restrict our weight parameter to be between 0 and 1, so in this case we shall put more attention on the median of the simulation results rather than the mean. The feature of the simulated distribution indicates that the FRW method can hardly improve forecasting accuracy. Such finding is consistent with Proposition 2 in Section 4 which shows the weights converge in probability to unity when the model is well specified. The right panel of Figure 5 shows the respective average sample weight and its 95% confidence bounds.

B.2 Experiment 2: Time-varying AR(1) model

In the second experiment, the time series are generated by an AR(1) model with time-varying coefficient, β, and the FRW parameter, ρ_1, is estimated based on Equation (10). The data generating process (DGP) is specified by:

$$\text{DGP: } y_t = \alpha + \beta_t y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{12}$$

$$\beta_t = 0.5 + 0.5\sin(2\pi t/B), \tag{13}$$

where $\alpha = 0.13$ and ϵ_ts are i.i.d distributed with variance $\sigma_\epsilon^2 = 0.5$. The coefficient, β is varying between 0 and 1 with respect to business cycle length $B = 72$. The exponential weight function is considered when applying FRW algorithm and we use Equation (10) with the estimates, $\hat{\boldsymbol{\theta}}_T(\mathbf{w}_k(\hat{\boldsymbol{\rho}_H}))$, from the FRW to calculate the forecasts.

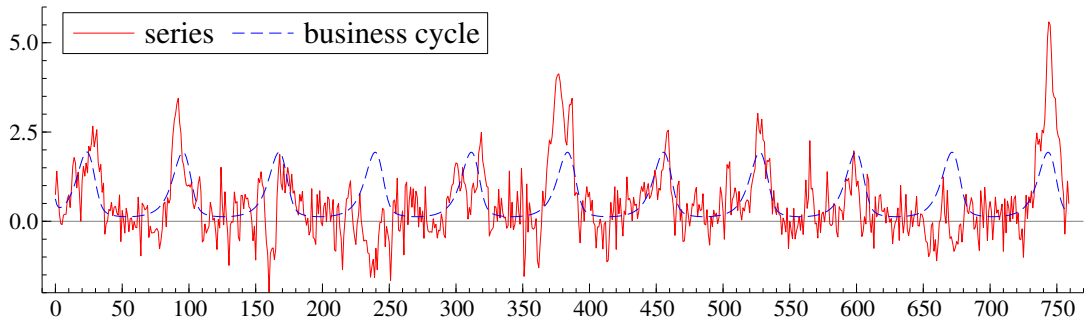


Figure 6: Illustration of DGP for Experiment 2.

In Experiment 2, we generate a time series of size $T = 760$. The generated data are considered as monthly data. We consider a DGP which includes a six-year business cycle, which is the average U.S. business cycle length. The rolling window size is set to be 60 which is a bit shorter than the business cycle. An illustration of such DGP is shown in Figure 6. Forecasts are made for time period $t = 701 \ldots 760$. In this experiment, the time series imply instability of the coefficients in the model and the original AR(1) model is misspecified, thus the FRW parameter ρ_1 is expected to be smaller than 1. This means recent observations are more relevant to the forecasts in the future.

The left panel in Figure 7 presents the simulated density of FRW parameter, ρ_1. The right panel shows the average sample weights and its 90% confidence bound. The right panel of Figure 7 shows that both the mean and the median of the simulated FRW parameter, ρ_1, is smaller than 1. The general picture of Figure 7 is that when the time series contains certain time-varying components in it and the considered forecasting model is misspecified,

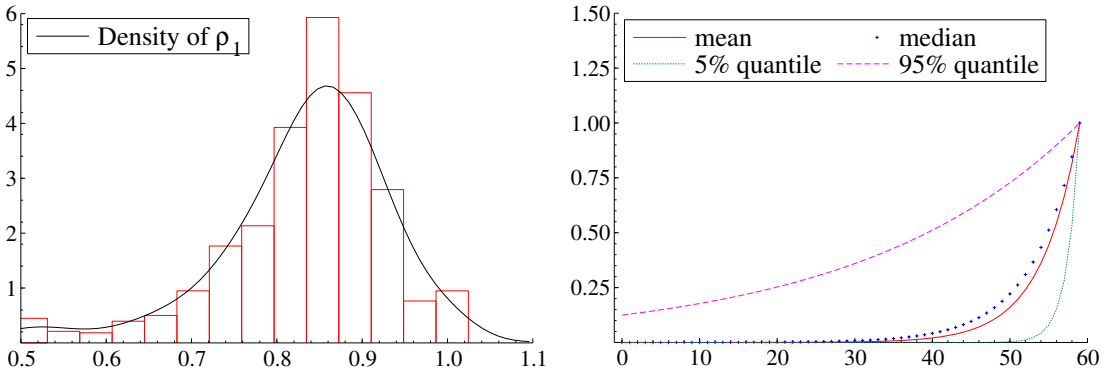


Figure 7: Simulation results of the FRW method over 1,000 simulations. The data are generated by an AR(1) model with the time-varying coefficient, β, and the FRW parameter, ρ_1, is estimated based on Equation (10). The exponential weight function is applied. The left panel presents the simulated density of the FRW parameter, ρ_1. The right panel shows the average sample weights and its 90% confidence bound.

the MLE method can improve forecasting accuracy by putting more weights to the recent observation. Such finding is also consistent with Proposition 3 in Section 4.

B.3 Experiment 3: AR(1) Model with a Structural Break

In Experiment 3, the data are generated by an AR(1) model with a structural break in the coefficient, β, and the FRW parameter, ρ_1, is estimated based on Equation (10). The data generating process (DGP) is specified by:

$$y_t = \alpha + \beta_t y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{14}$$

$$\beta_t = 0.2 + 0.7 I_t, \tag{15}$$

where $\alpha = 0.13$ and ϵ_ts are i.i.d normally distributed with variance $\sigma_\epsilon^2 = 0.5$. The indicator I_t is set to $I_t = 0$ for $t < 420$ and $I_t = 1$ for $t \geq 420$. The exponential weight function is considered when applying FRW algorithm and we use Equation (10) with the estimates, $\hat{\boldsymbol{\theta}}_T(\mathbf{w}_k(\hat{\boldsymbol{\rho}_H}))$, from the FRW to calculate the forecasts.

In Experiment 3, we generate a time series with size $T = 760$. The rolling window

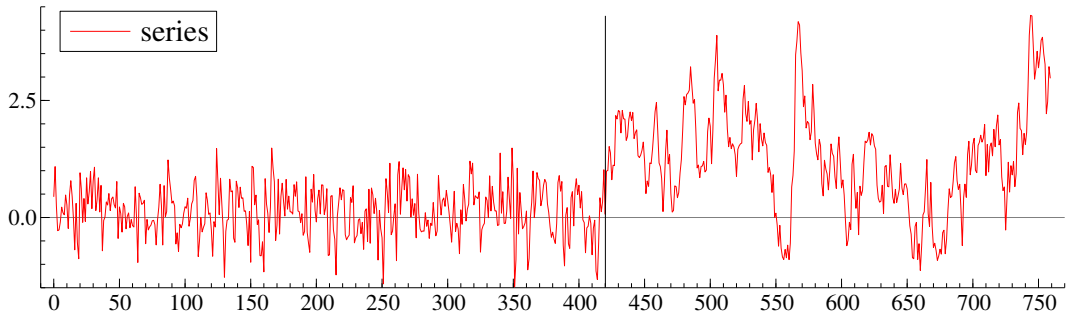


Figure 8: Illustration of DGP for Experiment 3.

length, τ, is selected to be 120. The break-point is set to $t = 420$. We evaluate the forecasting performance in three different forecasting periods. The first forecasting period is the period before the break-point (BB) starting from $t = 360$. The second one is the period right after the break-point (JAB) starting from $t = 444$ and the last one is the period long after (LAB) the break-point starting from $t = 492$. For all cases, we evaluate the forecasting performances of 60 observations. The generated data are considered as monthly data. This means, the length of forecasting period is 5 years. Figure 8 presents a realization of the data from Experiment 3. When forecasts are calculated before the break point, the time series is generated by an ordinary AR(1) model and the FRW converges to the classical ML method, while the simulated FRW parameter is expected to be 1. When forecasts are calculated just after the break point, the data before the break point are less relevant for calculating future forecasts, thus the FRW parameter, ρ_1, will be significantly smaller than 1. As the forecasting point is getting far away from the break point, the FRW parameter ρ_1 moves generally back to 1 again.

Figure 9 presents the density of FRW parameter, ρ_1, in Experiment 3. The FRW method can hardly improve forecasting accuracy for the forecasting period before the breaking point because the forecasting model is well specified just as the finding of Experiment 1. For the forecasting period right after the structural break the simulated mean of the FRW

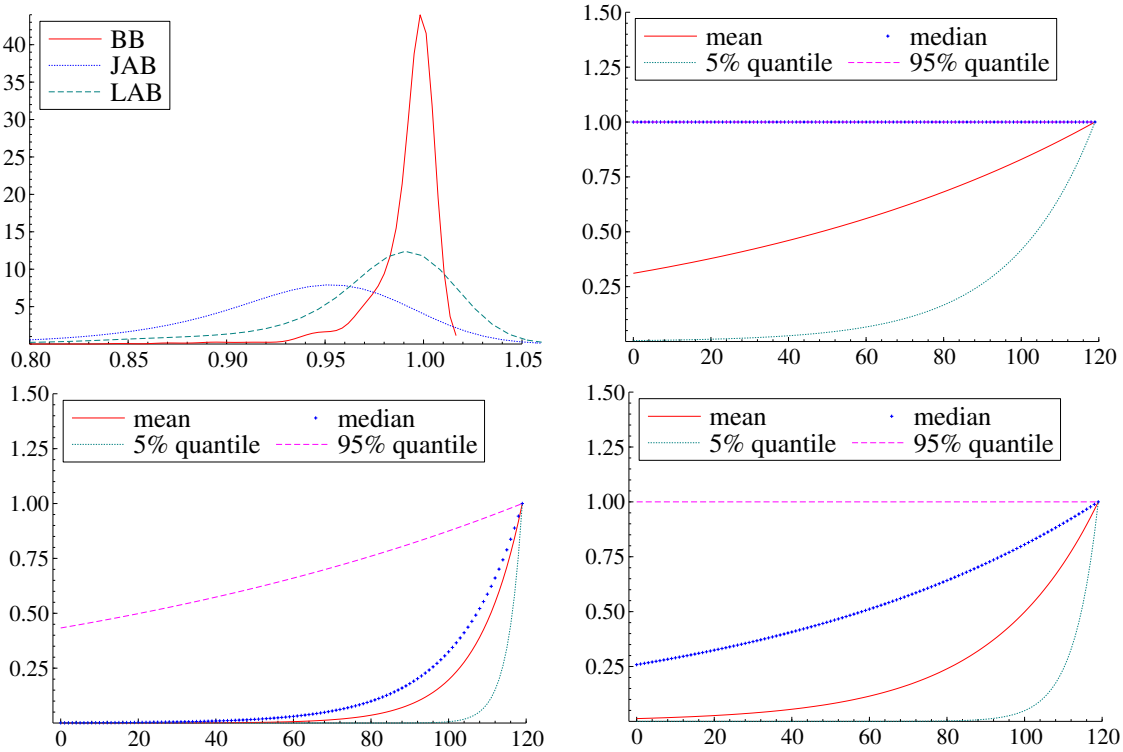


Figure 9: Simulation densities of FRW parameter, ρ_1, over 1,000 simulations. The data are generated by an AR(1) model with a structural break in the coefficient, β, and the FRW parameter, ρ_1, is estimated based on Equation (10). Exponential weight function is applied. The upper-left panel presents the simulated densities for three forecasting periods: the period before the breaking point (BB), the period just after the breaking point (JAB) and the period long after the breaking point (LAB). The upper-right panel presents the average sample weights and its 90% bound for forecasting period before the breaking point. The bottom-left panel presents the average sample weights and its 90% bound for forecasting period right after the breaking point. The bottom-right panel presents the average sample weights and its 90% bound for forecasting period long after the breaking point.

parameter is 0.93485, which indicates more weight should be put on recent observation in order to provide better forecasts. Finally, for the period long after structural break, fewer observations before the breaking point are included in the estimation window and the FRW parameter tends to peak at 1 again.

C Further empirical findings

Given the considerable delay in the NBER recession publications, as a robustness check, we consider a more timely recession indicator. Specifically, we consider the following recession weights $w_t = \rho_2 \cdot (1 - \hat{Z}_t) + (1 - \rho_2) \cdot \hat{Z}_t$, where $\hat{Z}_t \equiv \mathbb{E}[Z_t | \mathcal{F}_{t-n}]$ is the n-step-ahead predicted recession regime given the past information. For the in-sample and cross-validation parts, we utilize the lags of the NBER recession indicator, i.e. $\mathbb{E}[Z_t | \mathcal{F}_{t-n}] = Z_{t-n}^{NBER}$, which are available ex-post and provide a precise measure of the business cycle regime. For the out-of-sample part, we construct a more timely recession indicator based on the lags of the 'Sahm rule' indicator, which is available at the time of forecasting. The Sahm rule indicator, recently introduced in the literature to signal recession onsets (Sahm, 2019), utilizes unemployment rate data and signals the start of the recession when the three-month moving average of the unemployment rate increases by 0.50 percentage points or more relative to the minimum of the three-month averages from the previous 12 months[1]. Hence, we construct a simple recession indicator where we define periods with the Sahm indicator exceeding 0.50 and with its positive growth rate as recession periods. It is important to note that this approach may potentially misclassify recession periods as expansions, and vice versa.

In our analysis, we proceed to forecast monthly U.S. IPI growth rate during the great recession, as outlined in Section 6.1. The forecasting results are presented in

[1] https://fred.stlouisfed.org/series/SAHMCURRENT.

Table 5. Overall, the results align comparably with the results in Table 1 with FRW notably outperforming ML in the majority of cases. This highlights the superiority of the FRW approach, even when the weights are constructed based on the 'predicted' regimes. Additionally, we note that we utilize a basic recession indicator, which could potentially benefit from further refinement. For instance, the threshold of 0.50 proposed in Sahm (2019) could be optimized further, as suggested in Feng and Sun (2020). Alternatively, a more complex model for predicting business cycle regimes could be developed. This could be particularly relevant for the n-step-ahead forecasts, where Z_{t-n}^{NBER} ($Z_{t-n}^{SahmRule}$) is clearly not the most accurate predictor of recessions at period t, leading to notably deteriorated results for the U.S. IPI growth rate 3-step-ahead forecasts. In Tables 6 and 7, we present the forecasting results for the U.S. Total Non-farm Payrolls and Unemployment rate based on the predicted business cycle regimes.

Table 5: Pseudo out-of-sample forecasts: Binary predicted business bycle regime weights; Industrial Production

	RMSE						MAE					
	n=1			n=3			n=1			n=3		
	Ratio (GW)	ρ_2^r	ρ_2^e	Ratio (GW)	ρ_2^r	ρ_2^e	Ratio (GW)	ρ_2^r	ρ_2^e	Ratio (GW)	ρ_2^r	ρ_2^e
AR(1)												
25 years	1.094	3.5	78.2	0.909**	3.4	19.2	1.159	1.4	54.2	0.980	0.5	13.1
	(0.296)			(0.010)			(0.300)			(0.296)		
50 years	0.926**	4.5	73.1	0.887***	5.2	21.6	0.886***	0.6	60.4	0.872***	1.7	10.9
	(0.021)			(0.001)			(0.000)			(0.000)		
AR(AIC)												
25 years	1.213	3.3	70.0	0.984	3.4	19.2	1.393	0.9	60.1	1.168*	0.5	13.1
	(0.153)			(0.386)			(0.139)			(0.094)		
50 years	0.938**	5.3	62.0	0.891***	5.2	21.6	0.927***	0.9	57.6	0.880***	1.7	10.9
	(0.049)			(0.000)			(0.005)			(0.000)		
ADL(1,1)												
25 years	1.009	5.0	64.4	1.975	4.0	39.9	1.034	3.8	52.3	3.446*	0.3	30.5
	(0.461)			(0.114)			(0.456)			(0.092)		
50 years	0.860**	5.9	60.7	0.925*	6.5	31.3	0.825***	1.1	50.3	0.931***	12.5	40.2
	(0.017)			(0.078)			(0.003)			(0.003)		
ADL(AIC)												
25 years	1.432	6.9	56.4	1.146*	4.1	27.9	1.564*	4.3	62.8	1.074	6.6	36.8
	(0.100)			(0.079)			(0.082)			(0.242)		
50 years	0.786*	4.4	52.8	0.933*	7.6	29.8	0.769*	0.4	53.4	0.947	15.3	44.2
	(0.058)			(0.084)			(0.052)			(0.201)		

Table 5: *FRW* forecasting results for monthly U.S. IPI growth rate. The column labeled Ratio shows the RMSE or MAE using the FRW method applied to AR/ADL models relative to the RMSE or MAE using the MLE method. The weights for the FRW are defined as $w_t = \rho_2^i \cdot (1 - \hat{Z}_t) + (1 - \rho_2^i) \cdot \hat{Z}_t$, where $\hat{Z}_t$ is the 'predicted' recession regime and $i \in \{r, e\}$ indicates whether recession or expansion period was used for cross validation. The columns labelled 'ρ_r' and 'ρ_e' contain the optimal features of the FRW for the binary weighting, multiplied by 100. The estimation of the FRW features is performed on the cross-validation sample once. The cross-validation sample is either recession or expansion period of 2000–2007. The out-of-sample forecasts are computed for 2008–2009 June (18 months).

We provide further (rolling-window) forecasting results for our empirical illustrations on the basis of different weight functions within the FRW procedure. Tables 8 and 9 present the forecasting results for two expansion periods for the monthly time series of U.S. IPI growth rate obtained from autoregressive models with binary and decaying weights for FRW. Tables 10 and 11 present the forecasting results for the monthly time series of U.S. Unemployment Rate and U.S. Total Non-farm Payrolls, respectively.

Table 6: Pseudo out-of-sample forecasts: Binary predicted business cycle regime weights; U.S. Total Non-farm Payrolls

	RMSE						MAE					
	n=1			n=3			n=1			n=3		
	Ratio (GW)	ρ_2^r	ρ_2^e	Ratio (GW)	ρ_2^r	ρ_2^e	Ratio (GW)	ρ_2^r	ρ_2^e	Ratio (GW)	ρ_2^r	ρ_2^e
AR(1)												
25 years	0.616***	0.6	23.2	0.983	0.1	12.4	0.623***	1.0	33.6	1.048	0.0	18.9
	(0.000)			(0.432)			(0.000)			(0.364)		
50 years	0.536***	1.5	22.1	0.880	0.0	17.8	0.541***	1.1	25.4	0.821***	1.2	17.4
	(0.000)			(0.104)			(0.000)			(0.001)		
AR(AIC)												
25 years	0.835	2.4	18.1	1.134	0.1	13.3	0.848**	0.8	39.3	1.052	0.4	15.3
	(0.122)			(0.185)			(0.026)			(0.458)		
50 years	0.733***	2.8	23.6	1.028	0.0	17.9	0.750***	1.4	24.6	1.032	0.0	22.9
	(0.002)			(0.426)			(0.000)			(0.404)		

Table 6: *FRW* forecasting results for monthly U.S. Total Non-farm Payrolls. For further details, we refer to Table 5.

Table 7: Pseudo out-of-sample forecasts: Binary predicted business cycle regime weights; U.S. Unemployment Rate

	RMSE						MAE					
	n=1			n=3			n=1			n=3		
	Ratio (GW)	ρ_2^r	ρ_2^e	Ratio (GW)	ρ_2^r	ρ_2^e	Ratio (GW)	ρ_2^r	ρ_2^e	Ratio (GW)	ρ_2^r	ρ_2^e
AR(1)												
25 years	0.644**	0.9	84.0	0.649***	1.0	83.4	0.621**	0.0	54.2	0.622***	0.4	57.8
	(0.020)			(0.006)			(0.010)			(0.000)		
50 years	0.679***	5.8	65.4	0.670***	5.8	71.2	0.637***	2.5	58.8	0.631***	3.5	90.2
	(0.008)			(0.007)			(0.001)			(0.000)		
AR(AIC)												
25 years	0.695**	1.9	86.1	0.718***	1.0	100	0.709**	3.2	81.6	0.724***	0.7	100.0
	(0.024)			(0.006)			(0.034)			(0.000)		
50 years	0.720**	6.6	90.4	0.736***	5.0	91.9	0.715***	1.2	89.4	0.732***	3.1	99.0
	(0.018)			(0.008)			(0.006)			(0.000)		

Table 7: *FRW* forecasting results for monthly U.S. Unemployment Rate. For further details, we refer to Table 5.

Table 8: AR model forecasting results: Binary weight with decaying, expansion period 2002Jan-2007Dec

		RMSE						MAE					
		n=1			n=3			n=1			n=3		
	window	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2
AR(1)													
FRW/MLE	25 years	0.982	0.934	100.0	0.979	1.000	100.0	0.984	0.921	92.2	0.975	1.000	100.0
		(0.328)			(0.112)			(0.359)			(0.101)		
	50 years	0.891**	0.934	100.0	0.999	1.000	60.1	0.940*	0.996	93.2	0.999	1.000	60.4
		(0.027)			(0.441)			(0.051)			(0.459)		
FRW/ORW	50 years	0.958*	48		0.993	435		1.004	48		0.993	426	
		(0.096)			(0.310)			(0.469)			(0.352)		
AR(AIC)													
FRW/MLE	25 years	0.969*	1.000	100.0	0.976*	1.000	100.0	0.970*	0.999	99.9	0.972*	1.000	100.0
		(0.061)			(0.083)			(0.050)			(0.059)		
	50 years	0.931**	0.988	100.0	0.999	1.000	60.1	0.908**	0.989	76.4	0.999	1.000	60.4
		(0.036)			(0.441)			(0.010)			(0.459)		
FRW/ORW	50 years	0.968**	162		0.993	435		0.987	174		0.993	426	
		(0.034)			(0.310)			(0.103)			(0.352)		

Table 8: Rolling-window forecasting results for monthly U.S. IPI growth rate. Entries Ratio show the root mean-squared-forecast-error (RMSE) or mean-absolute-error(MAE) using FRW method of the AR model relative to RMSE or MAE using ML or ORW. Columns labelled ρ_1 and ρ_2 show the optimal weight parameter for the combined weight function. The forecasts were computed over the sample period 2002Jan–2007Dec. For further explanations, we refer to Table 1.

Table 8: AR model forecasting results: Binary weight with decaying, expansion period 2009Jul-2023Dec.

		RMSE						MAE					
		n=1			n=3			n=1			n=3		
	window	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2	Ratio	ρ_1	ρ_2
AR(1)													
FRW/MLE	25 years	0.751	0.976	100.0	0.949	0.974	100.0	0.881	0.955	100.0	1.046	0.966	100.0
		(0.155)			(0.298)			(0.114)			(0.195)		
	50 years	0.804	0.976	100.0	1.033	0.974	100.0	0.878*	0.955	100.0	1.078*	0.966	100.0
		(0.148)			(0.297)			(0.069)			(0.097)		
FRW/ORW	50 years	0.471	126		1.017	567		0.762	105		1.054	351	
		(0.154)			(0.425)			(0.120)			(0.167)		
AR(AIC)													
FRW/MLE	25 years	0.762	0.995	100.0	0.899	0.974	100.0	0.933	0.986	100.0	1.016	0.966	100.0
		(0.128)			(0.208)			(0.230)			(0.400)		
	50 years	0.854	0.990	100.0	1.033	0.974	100.0	0.974	0.983	100.0	1.078*	0.966	100.0
		(0.155)			(0.297)			(0.373)			(0.097)		
FRW/ORW	50 years	0.744	258		1.017	567		0.917	240		1.054	351	
		(0.135)			(0.425)			(0.199)			(0.167)		

Table 9: Rolling-window forecasting results for monthly U.S. IPI growth rate. Entries Ratio show the root mean-squared-forecast-error (RMSE) or mean-absolute-error(MAE) using FRW method of the AR model relative to RMSE or MAE using ML or ORW. Columns labelled ρ_1 and ρ_2 show the optimal weight parameter for the combined weight function. The forecasts were computed over the sample period 2009Jul–2023Dec. For further explanations, we refer to Table 1.

Table 10: AR model forecasting results: U.S. Unemployment Rate

		RMSE				MAE			
		n=1		n=3		n=1		n=3	
	Window	Ratio	ρ_2	Ratio	ρ_2	Ratio	ρ_2	Ratio	ρ_2
AR(1)									
FRW/MLE	25 years	0.608**	1.0	0.617***	2.0	0.639**	1.7	0.692***	5.0
		(0.019)		(0.007)		(0.028)		(0.001)	
FRW/MLE	50 years	0.673***	7.5	0.701***	10.2	0.782**	14.2	0.849***	24.6
		(0.005)		(0.003)		(0.010)		(0.000)	
FRW/ORW		0.671***	387	0.700***	378	0.781**	387	0.845***	378
		(0.005)		(0.003)		(0.011)		(0.000)	
AR(AIC)									
FRW/MLE	25 years	0.637**	1.7	0.691**	2.7	0.674*	2.2	0.706***	10.5
		(0.030)		(0.010)		(0.056)		(0.006)	
FRW/MLE	50 years	0.734**	9.4	0.750**	19.7	0.759**	8.7	0.848***	33.6
		(0.021)		(0.010)		(0.049)		(0.005)	
FRW/ORW		0.740**	576	0.746***	378	0.765*	576	0.844***	378
		(0.021)		(0.009)		(0.050)		(0.004)	

Table 10: Rolling-window forecasting results for monthly U.S. Unemployment. Entries Ratio show the root mean-squared-forecast-error(RMSE) or MAE using FRW method of the AR model relative to RMSE or MAE using normal ML or ORW methods. Columns labelled ρ_2 show the optimal weight parameter, multiplied by 100, for the combined weight function or optimal rolling window size. The forecasts were computed over the sample period 2008Jan–2009June.

Table 11: AR model forecasting results: U.S. Total Non-farm Payrolls

		RMSE				MAE			
		n=1		n=3		n=1		n=3	
	Window	Ratio	ρ_2	Ratio	ρ_2	Ratio	ρ_2	Ratio	ρ_2
AR(1)									
FRW/MLE	25 years	0.623***	0.1	0.734*	0.6	0.423***	1.9	0.622**	3.0
		(0.000)		(0.078)		(0.000)		(0.028)	
FRW/MLE	50 years	0.519***	2.5	0.553***	3.1	0.460***	5.1	0.497***	5.8
		(0.000)		(0.000)		(0.000)		(0.000)	
FRW/ORW		0.956	135	0.795***	132	0.887	135	0.728***	132
		(0.368)		(0.003)		(0.256)		(0.004)	
AR(AIC)									
FRW/MLE	25 years	0.755**	0.6	0.855	0.7	0.584***	1.9	0.715*	2.1
		(0.033)		(0.236)		(0.004)		(0.091)	
FRW/MLE	50 years	0.614***	2.8	0.564***	3.5	0.513***	8.1	0.512***	6.3
		(0.002)		(0.002)		(0.000)		(0.000)	
FRW/ORW		0.840**	135	0.593***	48	0.794**	135	0.526***	48
		(0.022)		(0.007)		(0.037)		(0.001)	

Table 11: Rolling-window forecasting results for monthly U.S. Total Non-farm Payrolls. Entries Ratio show the root mean-squared-forecast-error (RMSE) or MAE using FRW method of the AR model relative to RMSE or MAE using normal ML or ORW methods. Columns labelled ρ_2 show the optimal weight parameter, multiplied by 100, for the combined weight function or optimal rolling window size. The forecasts were computed over the sample period 2008Jan-2009June.