Erasmus University Rotterdam

MSc Econometrics and Management Science

Quantitative Finance

Master Thesis Final

Rolling window selection for out of sample forecasting in panel data models with time varying parameters

Supervisor
Wendun Wang

Author
Marcel Jongsma
387268

Co-Reader
Karolina Scholtus

Date:
October 2, 2019

Abstract

This thesis proposes a novel way to handle the presence of time varying parameters in economic time series by applying the method of Inoue et al. (2016), a method that finds an optimal window of observations to include when performing out of sample forecasts, on panel data models. In order to make the newly created combination more widely applicable a dynamic cluster method is introduced, which clusters units while accounting for the time varying property of the individual units in panel data. Performance is measured using the Root Mean Squared Forecast Error (RMSFE) on a data set of fifteen OECD countries for forecasting GDP growth. RMSFE is improved from 0.738 to 0.427 when introducing dynamic clusters and the optimal moving window method of Inoue et al. (2016). As results show slight improvements compared to other methods, this paper indicates a potential benefit of combining optimal moving window methods with panel data models.

Contents

1 Introduction

'The world is changing faster than ever'; a statement that you may have heard before. There are people that argue in favor of this, yet also people that argue against it. Whether or not it is changing faster than ever, the fact is that the world is changing. The world we know today is very different from the world we knew 10 years ago. Therefore, the point of this paper: do the dynamics of macroeconomic or financial models change over time as well? And if so, how can we deal with this?

The question at hand can be translated to whether there is parameter instability in macroeconomic and financial models. Substantial research has been conducted on this subject. J. Stock & Watson (1996), for example, conclude that parameter instability is present in a significant fraction of US macroeconomic time-series. Paye & Timmermann (2005) observe instability in parameters of linear models on international equity indices. Rossi (2013) and Clements & Hendry (1998) constitute prominent examples acknowledging the presence of time-varying parameters, while still leaving unmentioned other examples. The large amount of examples shows that parameter instability poses a challenge to many forecasting models.

Given that many financial and macroeconomic models consist of time-varying parameters, what implications does this have? Assume a situation where parameters are time-varying and estimations are based on historical data. Historical data, corresponding with different parameter values, behaves different compared to data that is currently generated, because theoretically, it corresponds with the current true parameter values. Therefore, by including historical data, a potential bias is introduced in the estimates. Consequently using these biased estimates to perform forecasts will make them inaccurate.

To improve forecasts, and hence reduce the bias in the parameter estimates, the general solution is to only include a limited number of recent observations when estimating the model. By doing so, the parameter estimates are based on data corresponding most accurately with the current true model. However, by including less data points, the variance of the estimates will increase. This leads to a trade-off between bias and variance in estimating a model's parameters. Estimating the model using a fixed number of observations and subsequently forecasting using such a model is called rolling out-of-sample forecasting, here the amount of data points included in estimating the model can be seen as a moving-window. The main issue in rolling out-of-sample forecasting is selecting the optimal moving window.

The selection of the optimal window can be done arbitrarily by forecasters, as for example done in J. H. Stock & Watson (2007), where inflation is forecasted using a 10-year window. Other methods designs attempt to find discrete breaks and choose the window length subject to the breaks. Pesaran & Timmermann (2007) propose five methods to find the optimal window such a situation. Finally, an optimal window can be selected under continuous breaks, as performed in Pesaran et al. (2011). A drawback of this method is the requirement of strictly exogenous regressors, which makes the method impossible to use in most time series. In my thesis I will deal with the latter situation leading to the method used in this paper, which is the method proposed by Inoue et al. (2016). In comparison to previous methods, this method has multiple advantages, most notably that the regressors can be weakly dependent and include both exogenous and lagged dependent variables.

The method of Inoue et al. (2016) is based on minimizing the conditional mean square forecast error (MSFE). However, as the conditional MSFE is infeasible, an approximate conditional MSFE is constructed by replacing the unknown parameters in the conditional MSFE with local linear estimates. Inoue et al. (2016) show that choosing the optimal moving window based on the approximated criterion is asymptotically equivalent to choosing the optimal moving window based on the infeasible one.

When we apply this method on macroeconomic data, which might consist of quarterly or even yearly data points the eventual variance of the estimates (or bias if the method includes more data points) might still be substantial. However, in many situations when dealing with country specific macroeconomic data, it is possible to use a panel data model. Using a panel data model will increase the amount of observations included when estimating the model. This increase in the number of observation reduces the variance of the estimates. With a reduced variance a new trade-off can be made, potentially reducing the amount of time observations included in the model, resulting in parameter estimates that are based on more current data. Therefore, this paper will attempt to find a practical way to apply the method of Inoue et al. on panel data, which is the main goal of this thesis.

In the rest of this section potential benefits and drawbacks of using panel data will be elaborated on; Panel data is also known as cross-sectional time series data or longitudinal data, consisting of data of a number of individuals (units) over time. Such data could, for example, comprise of different households' income or countries' GDP over time. The basic form of a panel data model further used in this thesis is defined as:

$$y_{i,t+1} = \alpha + \beta x_{it} + \mu_i + v_{it}, i = 1, ..., N; t = 1, ..., T \tag{1}$$

Where i specifies the unit and t specifies time, μ_i denotes the unit specific effect and v_{it} the error term. This basic model is divided into a fixed effects model (where the individual specific effect μ_i is seen as a fixed parameter to be estimated) and a random effects model (where μ_i is seen as a random variable).

To further substantiate why applying the method of Inoue et al. (2016) on panel data might turn out to be beneficial. Mark & Sul (2011) provide a good example. Suppose the data generating process is defined as follows:

$$y_{i,t+1} = \beta x_{it} + v_{i,t+1}, i = 1, ..., N, t = 1, ..., T \tag{2}$$

Where $v_t = v_{1,t}, ..., v_{N,t} \overset{iid}{\sim} (0, \sigma_v^2)$. Heterogeneous individual effects are left out in the DGP. Now asymptotic distribution theory shows that there are advantages of performing pooled estimation of the model (panel data model) over single unit regressions. The convergence rate of the panel data model estimator is given by:

$$\sqrt{NT}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma_v^2 Q^{-1}) \tag{3}$$

where $Q = plim((NT)^{-1} \sum_i \sum_t x_{it}^2)$. Meaning that the panel data estimator converges in distribution at rate $\sqrt{NT}$, compared to the single unit estimator, which converges at rate $\sqrt{T}$. This shows that the variance of the estimator reduces with the amount of units that are included in the panel data model. Such a reduction is conditioned by the assumption that all units have similar DGP. Under the assumption of time varying parameters, this result of reduced variance in panel data models creates more room to reduce the amount of time observations in estimating the model while keeping the variance low. However, a new bias is introduced by incorrectly assuming similar slope coefficients among units.

Panel data models often assume homogeneity among slope parameters, even though this might empirically not hold. Estimates can become biased if parameter homogeneity is incorrectly assumed. However, if heterogeneity is allowed, the model possibly includes too many parameters and estimates are not consistent anymore. To reduce the bias introduced by the assumption of homogeneity, while keeping the model parsimonious, a clustering of slope parameters can be considered. Lin & Ng (2012) introduce two methods to create such clusters. The first approach concerns a two-step pseudo threshold approach, which creates thresholds a variable that clusters each individual as either above or below the threshold. The second approach encompasses conditional K-means clustering, which consists of an adjusted K-means algorithm to allow for covariates. Similarly Bonhomme & Manresa (2015) propose a method to find optimal clusters for the individual effects of panel data models, which is also based on a K-means algorithm. As Lin & Ng (2012) describe in their paper, the threshold approach requires the estimates of the individual units, which are $\sqrt{T}$ consistent, opposed to the K-means approach which uses the group specific estimates which are $\sqrt{N_g T}$ consistent. Moreover, they mention that the method is computationally faster, yet sacrifices accuracy. These results combined with the aim of this thesis to lower the T dimension substantiate the superiority of using K-means in this thesis, which is therefore the method used.

Where homogeneity is assumed in most panel data models slope parameters, the individual effect is usually heterogeneous for each unit. One of the most known problems in panel data models is the incidental parameters problem. As noted by Neyman & Scott (1948), noise in the estimation of individual level effects when the time dimension of the panel is short will in general contaminate estimates of the common parameters. Given the fact that potentially fewer observations are included to account for time varying parameters, this incidental parameters problem potentially jeopardises estimates when applying the method of Inoue et al. (2016) on a panel data model.

To deal with the incidental parameters problem, either the amount of time observations has to increase, or the amount of fixed effects parameters has to be decreased. The latter can be achieved by clustering the fixed effects parameters, or to assume the individual effect of each unit to be homogeneous. In both ways, however, it is likely a new bias is introduced in the estimates. In the scenario of creating clusters Bester & Hansen (2009) show that clustering individual effects creates a trade-off between two biases; namely the bias due to the incidental parameters problem and the bias created by model miss-specification. In the scenario of assuming all individual effects to be homogeneous a bias can arise from this potentially incorrect assumption alone. However, in the application of this thesis I decided not to include a clustering of individual effects. The incidental parameters problem is present when the amount of units is high and the amount of time observations is low, which is not the case in the empirical part of this thesis.

It might be argued that, when modelling macroeconomic time series, a regular panel data model might not yield optimal performance, rather, the model's fit might be improved by adding a lagged dependent variable. When adding such a lagged dependent variable, the model transforms to a dynamic panel data model. A known problem when estimating the coefficients of this dynamic model concerns inconsistent estimations for a low time dimension (T). It is shown by Nickell (1981) that for a simple AR(1) panel model without exogenous regressors there is a bias in the coefficient of the lagged dependent variable, which disappears as the time dimension T increases. In order to address this problem, several solutions have been proposed, of which two well-known include instrumental variables (IV) Anderson & Hsiao (1982) and generalized method of moments (GMM) Manuel Arellano (1991) estimators. These solutions are, however, subject to some drawbacks. Both the GMM and IV estimators have relatively large standard errors when compared to the normal biased estimator. Moreover, when making use of IV estimation, additional decisions regarding instrumental variable selection is required. In 1995, Kiviet (1995) proposes a new analytical solution to the bias of the normal least squares dummy variable (LSDV) estimator. Somewhat later Judson & Owen (1999) compare the performance of different estimation methods and conclude that the bias-correction method as proposed in Kiviet (1995) gives the best estimates for all lengths (time dimension) of panel data, given that the panel data is balanced. However, also this bias correction method has its drawbacks; it is subject to strict theoretical restrictions and its calculation requires the use of a preliminary consistent estimator. In more recent papers such as Bun & Kiviet (2003), Bruno (2005) and Bun & Carree (2005), more novel approaches are proposed to achieve an unbiased estimate. In my thesis, I make use of another novel approach, namely the bootstrap based solution as formulated in the paper of Everaert & Pozzi (2007). Using a bootstrap based approach the data is simulated multiple times using some input parameters, once the estimates from these generated data sets correspond with the estimates of the original data set, they conclude that the input parameters of the simulation must correspond with the 'true' parameters. The authors conclude that their results are similar to the analytical bias corrections in samples with small to moderate T and outperform the GMM estimators in samples with small to moderate T and small N. Because of the favorable results, less restrictive method and easy implementation I choose to include this method to account for the potential bias.

Previous literature has not yet combined the usage of an optimal moving window selection method on a panel data model. As shown with the example before, the estimates of parameters will converge to the true parameter more promptly in a panel data model compared to a regular time series. This result, together with the trade-off between bias and variance and the related observation that the variance will be lower when employing panel data, vouches for the potential benefit of combining the two methods. However, combining these two methods introduces a double trade-off. First, a trade-off between the bias arising from including many historical data points versus the variance by including less time data points. Second, a trade-off between a newly introduced bias following from wrongly assuming units to be homogeneous versus a higher variance by not assuming units to be homogeneous. In this thesis I strive to find a comprehensive way to combine these two trade-offs to be applied in economic time series forecasting.

The remainder of this paper is structured as follows. First, the methodology used in the research will be elaborated upon (chapter 2), covering the explanation of panel data models, optimal moving window methods, and how to combine them. Then, a simulation experiment is conducted (chapter 3), followed by an empirical experiment (chapter 4). Finally, the results of these experiments are discussed and concluded upon (chapter 5).

2 Methodology

In this section I will elaborate a bit more on Panel Data models and how to get proper estimates for a panel data model. Thereafter I will explain the method of Inoue et al. (2016), after which I will explain how the optimal moving window method can be applied on panel data models. Lastly, methods to find clusters in the panel data will be elaborated on.

2.1 Panel Data Models

In this section the models used in this thesis will be explained.
The property that distinguishes panel data is that the variables in panel data have double subscripts, namely:

$$y_{i,t+1} = \alpha + x'_{it}\beta + u_{it}, i = 1, ..., N; t = 1, ..., T \tag{4}$$

Where i denotes the unit and t denotes time, in the empirical application subscript i will denote country. Furthermore α is a scalar, β is $K \times 1$ and x_{it} is the ith observation on K explanatory variables. The error term u_{it} can be broken up into μ_i and v_{it}, where μ_i denotes the unobservable individual specific effect (country specific effect in the empirical application) and v_{it} is the remainder disturbance.

It is possible to rewrite this model into vector form. First define $D = I_N \otimes \iota_T$. D is a selector matrix of ones and zeros, it can be seen as a matrix of individual dummies that correspond with the units in the panel data set. After that stack the observations over time to get: $y_{i,+1} = (y_{i2}, ..., y_{iT+1})', x_i = (x_{i1}, ..., x_{iT})', v_i = (v_{i1}, .., v_{iT})'$ and consequently stack over cross sections $y = (y_{1,+1}, ..., y_{N,+1})', x = (x_1, ..., x_N)', \mu = (\mu_1, ..., \mu_N)', v = (v_1, .., v_N)'$. Then using this and D results in:

$$y_{+1} = \alpha\iota_{NT} + x\beta + D\mu + v \tag{5}$$

Here y is $NT \times 1$, α is a scalar, x is $NT \times K$, β is $K \times 1$, D is $NT \times N$, μ is $N \times 1$ and v is $NT \times 1$.

There are two main sorts of panel data models, namely the fixed effects model and the random effects model. In the random effects model the individual parameter μ_i is assumed to be random according to a certain distribution. In the fixed effects model μ_i is seen as fixed parameter that has to be estimated. As mentioned in Baltagi (2012) the fixed effects model is an appropriate specification when focusing on a specific set of N units, as for instance firms, countries or states. Since the models will be applied on economic time series, the fixed effects model is more suitable and will be used in the rest of my thesis.

By choosing to use the fixed effects model the parameters μ_i is fixed and has to be estimated, the remaining disturbance v_{it} is independent and identically distributed $IID(0, \sigma_v^2)$ and the x_{it} are assumed independent of v_{it} for all i and t.

Now either one μ_i can be set to zero, or α and μ are combined in order to make the model identifiable. In this thesis the latter option is chosen as it is easier to work with. Therefore, the parameter μ_i will actually be $\mu_i + \alpha$ and the part $\alpha\iota_{NT}$ will be removed from the model. So what is left is:

$$y_{+1} = x\beta + D\mu + v \tag{6}$$

This equation can be estimated with simple OLS. However, as one can see that if N is large, the model will include too many individual dummies. Therefore, the matrix to be inverted by OLS explodes. The most known way to solve this is by making use of the least squares dummy variable (LSDV) estimator (also known as the within estimator). When using the LSDV estimator, the variables in the model are demeaned in order to remove the individual effect. This is done by pre-multiplying the model with matrix Q, which is defined as: $Q = I_N(I_T - 1/T\iota_T\iota_T')$. Q is a symmetric idempotent matrix, which means that rank(Q) = tr(Q) = N(T-1) and $Q = Q'$ and $Q^2 = Q$. After demeaning simple OLS is used on the resulting transformed model:

$$Qy_{+1} = Qx\beta + Qv \tag{7}$$

Note that $QD = 0$ and that the elements of Qx and Qy are demeaned in the following way: $x_{it} - \bar{x}_i$ and $y_{it} - \bar{y}_i$.
The LSDV estimator $\tilde{\beta}$ thus comes down to:

$$\tilde{\beta} = (x'Qx)^{-1}x'Qy_{+1} \tag{8}$$

As argued before, modeling macroeconomic data with a simple panel data model might not give the best results. A dynamic panel data model might perform better. Here the extra term $y_{i,t-1}$ is added in the model to result in the following.

$$y_{i,t+1} = y_{i,t}\gamma + x_{i,t}'\beta + \mu_i v_{i,t} \tag{9}$$

Which in vector form translates to:

$$y_{+1} = y'\gamma + x'\beta + D\mu + v \tag{10}$$

As Nickell (1981) has shown performing LSDV or OLS on the dynamic panel data model gives biased estimates. Unless the intercept is completely omitted the estimates will be biased, but leaving out the intercept will most likely worsen the fit of the model. Therefore, a bias correction has to be included, which is explained in the next section.

2.2 Bootstrap-based bias correction

In this section a short summary of the bootstrap method of Everaert & Pozzi (2007) will be given, which corrects the biased estimates using the LSDV estimator in dynamic panel data models. The dynamic panel model is given as:

$$\begin{aligned} y_{+1} &= y'\gamma + x'\beta + D\mu + v \\ &= W\delta + D\mu + v \end{aligned} \tag{11}$$

γ and β are clustered into parameter vector δ and y and x are clustered in observation vector W. Now define the LSDV estimator $\hat{\delta}$ as:

$$\hat{\delta} = (W'QW)^{-1}W'Qy_{+1} \tag{12}$$

Recall the Q matrix $Q = I_N(I_T - 1/(T)\iota_T\iota_T')$ Given the results of Nickell (1981) it is evident that $\hat{\delta}$ is a biased estimator for δ. In order to come to unbiased estimates, the idea of Everaert & Pozzi (2007) boils down to the following: 'if we would sample repeatedly from a population with parameters $\bar{\delta}$ and calculate the LSDV estimate $\hat{\delta}_j^*(\bar{\delta})$ in each sample, $\bar{\delta}$ is an unbiased estimate for δ if the average of $\hat{\delta}_j^*(\bar{\delta})$ over J samples corresponds to the LSDV estimate $\hat{\delta}$ of δ based on the original data.'

The execution of this method is according to the following procedure:

After getting the first set of biased estimates $\hat{\delta}$ and for some input parameters $\tilde{\delta}$,

1. Estimate the individual effect $\tilde{\mu} = (T)^{-1}D'(y_{+1} - W\tilde{\delta})$ and the residuals $\tilde{v} = y_{+1} - W\tilde{\delta} - D\tilde{\mu}$.

2. Then for every bootstrap sample $j \in B$

(A) obtain a bootstrap sample $\tilde{v}^b$ from the rescaled estimated residuals $\tilde{v}^r$ from the following formula:

$$\tilde{v}_{it}^r = \sqrt{\frac{(T)}{(T-1)}} \left(\frac{\tilde{v}_{it}}{\sqrt{m_{it}}} - \frac{1}{(T)} \sum_{s=1}^{T} \frac{\tilde{v}_{is}}{\sqrt{m_{is}}} \right) \tag{13}$$

Here m_{it} is the itth diagonal element of the projection matrix $M = (I_{N(T)} - QW(W'QW)^{-1}W')Q$.

(B) Next, calculate a bootstrap sample $y_{+1}^b = W^b\tilde{\delta} + D\tilde{\mu} + \tilde{v}^b$ for the variable y. Where $W^b = (y^b, x)$, with the initialization $y_{i1}^b = y_{i1}$. Meaning that it is conditioned on the initial values y_{i1} and on x.

(C) Then obtain the LSDV estimator $\tilde{\delta}_j^b = (\tilde{\gamma}_j^b, \tilde{\beta}_j^b) = (W^{b\prime}QW^b)^{-1}W^{b\prime}Qy_{+1}^b$.

3. Over all B bootstrap samples calculate the mean of the LSDV estimator as: $\tilde{\delta}^b = B^{-1}\sum_{j=1}^{B}\tilde{\delta}_j^b(\tilde{\delta})$.

As mentioned before, in order for $\tilde{\delta}$ to be an unbiased estimator for δ, the mean of the bootstrap distribution $\tilde{\delta}^b$ should be equal to the original biased LSDV estimates $\hat{\delta}$. This translates to the following equation: $\omega = \hat{\delta} - \tilde{\delta}^b = 0$. If this equation is true then $\tilde{\delta}$ is an unbiased estimator of δ. In order to find the unbiased estimates, the bootstrap procedure is repeated several times until in an iteration k, $\omega_{(k)} = 0$ and $\tilde{\delta}_{(k)}$ is taken as an unbiased estimate for δ. As long as $\omega_{(k)} \neq 0$, $\tilde{\delta}_{(k)}$ is updated as $\tilde{\delta}_{(k+1)} = \tilde{\delta}_{(k)} + \omega_{(k)}$. The algorithm is initialized by setting $\tilde{\delta}_{(1)} = \hat{\delta}$. When creating a bootstrap sample, for every observation a random value from the adjusted residual is chosen with replacement.

2.3 Selecting optimal moving window

In this section I will explain the method of Inoue et al. (2016). The method will be simplified to one step ahead forecasts. Let us assume a data generating process (DGP) of:

$$y_{t+1} = \beta_t' x_t + u_{t+1}, t = 1, 2, ..., T, \tag{14}$$

With x_t a $K \times 1$ vector of stochastic regressors, β_t is a $K \times 1$ vector of time varying parameters and u_{t+1} an unobservable disturbance. T is the full sample size. It is important to note that x_t may include exogenous explanatory variables as well as lagged values of the dependent variable.

The method is based on using local linear estimates of the parameters. In order to be able to create a local linear estimate the parameter β_t has to be twice differentiable over time. This is done by representing β_t as a smooth function over current period t. Thus β_t is defined as: $\beta(\frac{t}{T})$. The parametric form of $\beta(\frac{t}{T})$ is unknown.

$$y_{t+1} = \beta\left(\frac{t}{T}\right)' x_t + u_{t+1} \tag{15}$$

$\beta(\frac{t}{T})$ is a vector of unknown smooth functions of time t.

The idea is that every point s "near" point t can be approximated using a Taylor expansion as follows:

$$\beta\left(\frac{s}{T}\right) = \beta(\frac{t}{T}) + \beta^{(1)}(\frac{t}{T})\left(\frac{s-t}{T}\right) + \frac{\beta^{(2)}(c)}{2!}\left(\frac{s-t}{T}\right)^2 \tag{16}$$

where $c = \lambda\frac{s}{T} + (1-\lambda)\frac{t}{T}$, for $\lambda \in (0,1)$ and $\beta^{(i)}$ is the ith derivative of $\beta(\frac{t}{T})$. The third term of the equation is negligible. Therefore the following holds:

$$y_{s+1} = \beta(\frac{t}{T})' x_s + \beta^{(1)}(\frac{t}{T})' x_s \left(\frac{s-t}{T}\right) + \epsilon_{s+1} \tag{17}$$

ϵ_{t+1} is the composite error term of u_{t+1} and the second order term in equation 16.

Now to forecast the next step at the end of the sample, $\frac{t}{T}$ is replaced with $\frac{T}{T}$, then the beta terms of equation 17 can be estimated using an OLS estimator as follows (note that $\frac{T}{T} = 1$):

$$\begin{bmatrix} \tilde{\beta}(1) \\ \tilde{\beta}^{(1)}(1) \end{bmatrix} = \begin{bmatrix} \sum x_s x_s' & \sum x_s x_s' \left(\frac{s-T}{T}\right) \\ \sum x_s x_s' \left(\frac{s-T}{T}\right) & \sum x_s x_s' \left(\frac{s-T}{T}\right)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_s y_{s+1} \\ \sum x_s y_{s+1} \left(\frac{s-T}{T}\right) \end{bmatrix} \tag{18}$$

The summations in the equation are $\sum\limits_{s=T-R_0+1}^{T-1}$, thus the local linear estimates are estimated using the most recent R_0 data, where R_0 is a given window size for the local linear regression. This window size is determined by the cross validation method of M. Hashem Pesarana (2007). This method will be explained in another section.

The local linear estimate is later on used in the estimation of the optimal moving window. The optimal moving window is chosen by minimizing the conditional Mean Squared Forecast Error (MSFE). The conditional MSFE is defined as follows:

$$E[(y_{T+1} - \hat{\beta}(1)'x_T)^2|\Omega_T] \tag{19}$$

Where $\hat{\beta}(1) = \hat{\beta}(\frac{T}{T})$ and Ω_T is the set of all information up to time T. Furthermore, since the goal is to minimize the conditional MSFE with respect to window size R $\beta(\frac{t}{T})$ is replaced with $\beta_R(\frac{t}{T})$ which is the estimate based on window size R defined as follows:

$$\hat{\beta}_R(\frac{T}{T}) = \left(\sum_{t=T-R+1}^{T-1} x_t x_t'\right)^{-1} \left(\sum_{t=T-R+1}^{T-1} x_t y_{t+1}\right) \tag{20}$$

Now the conditional MSFE can be rewritten.

$$\begin{aligned} E[(y_{T+1} - \beta\hat{(1)}'x_T)^2|\Omega_T] &= E[(\beta(1)'x_T + u_{T+1} - \beta\hat{(1)}'x_T)^2|\Omega_T] \\ &= E[u_{T+1}^2|\Omega_T] - 2E[(\hat{\beta_R}(1) - \beta(1))'x_T u_{T+1}|\Omega_T] \\ &+ E[(\hat{\beta_R}(1) - \beta(1))'x_T' x_T(\hat{\beta_R}(1) - \beta(1))|\Omega_T] \end{aligned} \tag{21}$$

Since $\hat{\beta_R}(1)$ and x_T are deterministic given information set Ω_T the second term will simplify to

$$E[(\hat{\beta_R}(1) - \beta(1))'x_T u_{T+1}|\Omega_T] = (\hat{\beta_R}(1) - \beta(1))'x_T E[u_{T+1}|\Omega_T] = 0 \tag{22}$$

As the expectation of the error term is equal to zero. And since the first term is the variance of the model, which is not dependent on moving window R. minimizing the conditional MSFE with respect to R is equivalent to minimizing the last term in equation 22. Thus minimizing the following:

$$E[(\hat{\beta_R}(1) - \beta(1))'x_T' x_T(\hat{\beta_R}(1) - \beta(1))|\Omega_T\} \tag{23}$$

However, since the actual value of $\beta(\frac{t}{T})$ is unknown it is replaced with the earlier defined local linear estimate. This can be done under the assumption that $\beta(\frac{t}{T})$ is twice differentiateble over t. This results in the approximate MSFE. Inoue et al. (2016) show that the approximate MSFE is asymptotically optimal relative to the infeasible conditional MSFE. So by plugging in the local linear estimate, the approximate MSFE is created, which will be minimized with respect to window size R, thus the optimal window size $\hat{R}$ satisfies:

$$\hat{R} = \underset{R \in \theta_R}{\arg\min} (\hat{\beta_R}(1) - \tilde{\beta}(1))'x_T x_T(\hat{\beta_R}(1) - \tilde{\beta}(1)) \tag{24}$$

2.4 Cross validation method to set initial window size

The input window size used in the window selection method is determined by the cross validation method as in M. Hashem Pesarana (2007). In this method the last $\tilde{\omega}$ observations are reserved for an out-of-sample estimation exercise and chooses the estimation window that generates the smallest MSFE value on this sample. Next to $\tilde{\omega}$ a minimum of $\underline{\omega}$ is needed to estimate the parameters of the model. In their paper they set $\underline{\omega}$ and $\tilde{\omega}$ to 25% and 10% of the data respectively, therefore, I choose to do the same.

The pseudo-out-of-sample MSFE is calculated for each starting point of the estimation window m. As follows:

$$MSFE(m|T,\tilde{\omega}) = \frac{1}{\tilde{\omega}} \sum_{\tau=T-\tilde{\omega}}^{T-1} (y_{\tau+1} - x_{\tau}^{'}\hat{\beta}_{m:\tau})^2 \tag{25}$$

Following this the optimal starting point m^* is the value of m where the sum given above is the smallest.

$$m^*(T,\underline{\omega},\tilde{\omega}) = \underset{m=1,\ldots,T-\underline{\omega}-\tilde{\omega}}{\arg\min} \left\{ \frac{1}{\tilde{\omega}} \sum_{\tau=T-\tilde{\omega}}^{T-1} (y_{\tau+1} - x_{\tau}^{'}\hat{\beta}_{m:\tau})^2 \right\} \tag{26}$$

Resulting in an input window of $R0 = T - m^*$.

2.5 Applying optimal moving window method on panel data model

In order the apply the method of Inoue et al. (2016) on panel data models, part of the method has to be rewritten to match with the LSDV estimator for panel data models. To start, the dynamic panel data model is written as a one step ahead forecast model:

$$y_{i,t+1} = y_{it}\gamma + x_{it}'\beta + \mu_i + v_i, \tag{27}$$

Which in vector form is:

$$\begin{aligned} y_{+1} &= y'\gamma + x'\beta + D\mu + v, \\ &= W\delta + D\mu + v, \end{aligned} \tag{28}$$

Now recall equation 17 which is rewritten slightly to fit the panel data model setting:

$$y_{i,s+1} = x_{is}'\beta(\frac{t}{T}) + x_{is}'\left(\frac{s-t}{T}\right)\beta^{(1)}(\frac{t}{T}) + \epsilon_{i,s+1}, \tag{29}$$

In the dynamic panel data model this translates to:

$$\begin{aligned} y_{i,s+1} &= w_{is}'\delta(\frac{t}{T}) + w_{is}'\left(\frac{s-t}{T}\right)\delta^{(1)}(\frac{t}{T}) + \epsilon_{i,s+1}, \\ &= \begin{bmatrix} w_{is}' & w_{is}'\left(\frac{s-t}{T}\right) \end{bmatrix} \begin{bmatrix} \delta(\frac{t}{T}) \\ \delta^{(1)}(\frac{t}{T}) \end{bmatrix} + \epsilon_{i,s+1}, \end{aligned} \tag{30}$$

The individual effect is left out. It is arbitrary to leave it out or include it. When controlling for parameter instability in the fixed effects the individual effect should be included, but since this would included N more variables (since the demeaning will not wipe out $\frac{s-t}{T}D$ from the equation) the inverse operation might become large and badly influence the estimation. Therefore, I chose to ignore the potential instability in the fixed effects and focus on the shared common parameters. Define $W_{is}^* = \begin{bmatrix} w_{is}' & w_{is}'\left(\frac{s-t}{T}\right) \end{bmatrix}'$ and $\delta^*(\frac{t}{T}) = \begin{bmatrix} \delta(\frac{t}{T})' & \delta^{(1)}(\frac{t}{T})' \end{bmatrix}'$. Then after stacking the observations W_{is}^* over time and units this equation can be written in vector form:

$$y_{+1} = W^{*\prime}\delta^*(\frac{t}{T}) + \epsilon \tag{31}$$

Now in order to come to the local linear estimate that is necessary, still the LSDV estimator is used to get the parameter estimates. Thus the equation is premultiplied by Q, with $Q = I_N(I_T - 1/(T)\iota_T\iota_T')$. Consequently solving the resulting equation will give the following parameter estimates:

$$\tilde{\delta}^* = (W^{*\prime}QW^*)^{-1}W'Qy_{+1}, \tag{32}$$

Since these estimates are still biased, the bootstrap method will be used to come to unbiased estimates. The piece of the parameter vector needed is the $\tilde{\delta}(\frac{T}{T})$, which is the first half of $\tilde{\delta}^*$

After defining the local linear estimate in the panel data model situation, the approximate MSFE has to be minimized.

$$\hat{R} = \underset{R\in\theta_R}{\arg\min} \sum_{i=1}^{N}[(\hat{\delta_R}(1) - \tilde{\delta}(1))'w_{iT}'w_{iT}(\hat{\delta_R}(1) - \tilde{\delta}(1))] \tag{33}$$

Here $\hat{\delta}$ is the bias adjusted estimate of the dynamic panel data model using R observations. The function is optimized over R to find optimal moving window $\hat{R}$ for a dynamic panel data model.

2.6 Cluster slope parameters

After defining the model used and the method for optimal moving window selection, the time series that can be clustered together have to be determined, in order to make the method implementable. In the case of the empirical section of my thesis this translates to: which countries share the same slope parameters and which do not.

To create a cluster of units the idea of this research has to be kept in mind, which is that the dynamics of macroeconomic variables are time-varying. This can translate into one units' dynamics changing over time differently than other units. Simply said, if there are two clusters one unit can change dynamics by moving from cluster 1 to cluster 2 over time, whereas the rest stays the same. Ideally, the clusters have to be determined by taking into account the time varying property of the parameters. Therefore, I will consider two ways of creating clusters. Firstly, by basing the clusters on a fixed window T and secondly, by finding the optimal window length for each cluster during the cluster algorithm. I will elaborate on the first method first and extend it afterwards.

To find optimal clusters a K-means like method will be used, similar to the methods proposed in Bonhomme & Manresa (2015) and Lin & Ng (2012). Adjusted to fit the model at hand.

Recall the standard fixed effects model with individual heterogeneity.

$$y_{i,t+1} = x'_{i,t}\beta + \mu_i + v_{it}, i = 1, ..., N; t = 1, ..., T \tag{34}$$

In order to create a cluster of units this model is rewritten to the following:

$$y_{i,t+1} = x'_{i,t}\beta_{g_i} + \mu_i + v_{it} \tag{35}$$

Where $g_i \in 1, ..., G$ denoting the cluster that individual i is in. γ is the set of g_i's $\in \Gamma_G$ with Γ_G the set of all possible clusters of individuals. Moreover $\beta \in B$ and $\mu \in U$.
Given the definitions the estimator is defined as the solution of the following minimization problem:

$$(\hat{\beta}, \hat{\mu}, \hat{\gamma}) = \underset{(\beta,\mu,\gamma)\in B\times U\times \Gamma_G}{\arg\min} \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t+1} - x'_{i,t}\beta_{g_i} - \mu_{g_i})^2 \tag{36}$$

To make this minimization problem computationally solvable it is divided in two steps, the first one minimizes the function with respect to cluster membership for given values of β and μ for each individual i. Which can be interpreted as choosing optimal cluster membership for each individual given the model parameters.

$$\hat{g}_i(\beta, \mu) = \underset{g\in 1,\ldots,G}{\arg\min} \sum_{t=1}^{T}(y_{i,t+1} - x'_{i,t}\beta_{g_i} - \mu_{g_i})^2 \tag{37}$$

The second one minimizes the function with respect to parameters β and μ, which can be interpreted as finding the optimal parameter estimates for given clusters $\hat{g}_i(\beta, \mu)$.

$$(\hat{\beta}, \hat{\mu}) = \underset{(\beta,\mu)\in B\times U}{\arg\min} \sum_{n=1}^{N}\sum_{t=1}^{T}(y_{i,t+1} - x'_{i,t}\beta_{g_{i(\beta,\mu)}} - \mu_{g_{i(\beta,\mu)}})^2 \tag{38}$$

To find correct clusters, random input parameters are given, however, some studies show that the best results are given when the input parameters are calculated by randomly creating a first cluster and estimate the model according to this cluster using the parameters of this model as the first set of input parameters. I will use this method as well. Optimal clusters are calculated according to the input parameters and consequently new parameters are calculated according to these new clusters. Then there are some iterations over these two steps. To make sure the global minimum is found rather than a local minimum, the method can be repeated several times with different starting values.

This method works for both the normal and dynamic panel data model, in the setting of the dynamic panel data model the x variable includes the lagged dependent variable. In the case of the dynamic panel data model, the parameters will be adjusted according to the bootstrap method to come to unbiased estimates.

In the scenario that the amount of clusters is unknown, it can be numerically estimated by using information criteria. As Lu & Huang (2011) explain in their paper, the Akaike (AIC), Bayesian (BIC) and HannanCQuinn (HQI) Information Criterion can be used to find optimal number of clusters. which are all defined as a combination of SSR and a penalty term for the amount of parameters used in the model. Where the BIC poses the biggest penalty on extra parameters, followed by the HQI and AIC. The information criteria applied to the setting of this paper are defined as follows:

$$AIC = 2(p \times G + N) + NT \times log(\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\mu}_i + x'_{i,t}\hat{\beta}_{g_i} - y_{i,t+1})^2) \tag{39}$$

$$BIC = (p \times G + N)log(NT) + NT \times log(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\mu}_i + x'_{i,t}\hat{\beta}_{g_i} - y_{i,t+1})^2) \tag{40}$$

$$HQI = 2(p \times G + N)log(log(NT)) + NT \times log(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\mu}_i + x'_{i,t}\hat{\beta}_{g_i} - y_{i,t+1})^2) \tag{41}$$

Here p is the amount of parameters for the independent variables and G the amount of groups. By calculating the Information Criteria for every amount of clusters and consequently selecting the amount of clusters that minimizes the Information Criteria value the optimal amount of clusters are found.

As mentioned before, this method uses a fixed window T to find the optimal clusters, whereas this might not be in line with the time varying property of this paper. Therefore, I will explain how I will include this property.

Similarly to before, the problem is divided in two steps. In the first step, the method will find optimal clusters given the estimates and optimal windows for each existing cluster. Yet basing the cluster selection on SSE will lead to bad results, as the smallest window length will most likely give the smallest SSE, therefore, to make the clusters comparable the MSE is used by dividing by each clusters corresponding optimal window. Consequently, the second step applies the method of Inoue et al. (2016) to find optimal windows and corresponding optimal estimates for each given cluster.

Until now, the adjustment to include the time varying property is quite straightforward, yet the tricky part is finding the optimal amount of clusters, as the AIC, BIC and HQI cannot directly be applied anymore, since they are based on SSE and T. These, possibly vary across amount of clusters and per cluster. Therefore, a new way to find optimal amount of clusters has to be found. I did not find any replacing method for amount of cluster selection and therefore I decided to create my own, based on making a small alteration to the existing AIC, BIC and HQI. The alteration is quite straightforward. Rather than using the SSE, the MSE times a 'selection value' (SV) will be used and this SV will thus also replace T in the equation. Now, various clusters will either be extrapolated towards the SV or shortened towards the SV to make them comparable in the information criteria. In the case of the AIC this would result in the following:

$$AICa = 2(p \times G + N) + N(SV) \times log(\sum_{i=1}^{N}(\sum_{t=1}^{R_{gi}}(\hat{\mu}_i + x'_{i,t}\hat{\beta}_{g_i} - y_{i,t+1})^2)/R_{gi} \times (SV)) \tag{42}$$

Here AICa stands for AIC altered. This way the different optimal lengths of each cluster (R_{gi}) are made equal so that they can be compared with each other.

In order to compare AIC and AICa with eachother let's look at a simple example:

$$AIC = n \times log(SSE) + 2k \tag{43}$$

Where n is the amount of observations and k is the amount of parameters. Basically what the AICa boils down to is to estimate the SSE using the MSE of a smaller or bigger sample. Thus $S\hat{S}E = MSE_m \times n$ where the MSE is calculated with m observations with $m \neq n$. SSE can be replaced with $S\hat{S}E + \epsilon$ where ϵ is an error term with unknown distribution. Resulting with:

$$AIC = n \times log(S\hat{S}E + \epsilon) + 2k \tag{44}$$

By taking the expectation a potential bias is checked for.

$$\begin{aligned} E(AIC) &= E(n \times log(S\hat{S}E + \epsilon) + 2k) && (45) \\ &= n \times E(log(S\hat{S}E + \epsilon)) + 2k && (46) \\ &= n \times \left(log(E(S\hat{S}E + \epsilon)) - \frac{V(S\hat{S}E + \epsilon)}{2E(S\hat{S}E + \epsilon)} \right) + 2k && (47) \\ &= n \times \left(log(S\hat{S}E) - \frac{V(\epsilon)}{2S\hat{S}E} \right) + 2k && (48) \end{aligned}$$

In this result it is assumed that the expectation of ϵ is zero. It is easy to see that the AICa is actually slightly biased, however, intuitively $V(\epsilon)$ would not be bigger than $S\hat{S}E$ and therefore, the bias is not bigger than $n \times 1/2$. In my paper I will assume this bias to be negligible and use the AICa to create clusters.

An important thing to note is that the input given to SV is very influential in the resulting value of the AICa. If SV is large, the criteria will put more weight on the MSE, rather than the amount of parameters added in the model. Yet if the SV is small, more weight is put on the amount of parameters, rather than the MSE. The actual choice of the SV is arbitrary, as a rule of thumb: the value should be smaller than the largest optimal window and bigger than the smallest optimal window (unless one wants to be extra lenient or strict).

3 Monte Carlo Experiment

Before conducting the Monte Carlo Experiment, keep in mind that the method will be used on an OECD data set with moderate to large T and small to moderate N depending on cluster of the countries. To test the method in a more realistic setting, the Monte Carlo simulation will be proposed such that it will be similar to the empirical data. Consequently, using a bootstrap bias correction in this scenario might not turn out as valuable. Table 13 in the appendix shows that the magnitude of the bias in the dynamic panel data models used in this paper does not demand the use of a bootstrap bias correction. Moreover, leaving out the bias correction will relieve the method with a considerable computational burden.

The method of combining the optimal moving window estimation and panel data models will be tested for both a dynamic and a static panel data model. Results will be compared with an auto-regressive model including exogenous variables and with a standard regressive model respectively. In order to do this, several Monte Carlo simulations will be performed, of which the data generating processes will be explained in the subsequent sections. A total of five experiments will be done, comprising of several DGPs.

3.1 Dynamic Panel Data Model

Similar to Inoue et al. (2016), who evaluate 22, mostly dynamic, DGPs, in this section, I will consider similar DGPs divided in three experiments in this section. The first experiment will analyze 8 DGPs, the second will analyze 6 DGPs, and the last will analyze a combination of the previous DGPs.

The DGPs of the first experiment follow, similarly to Inoue et al. (2016), the following process:

$$\begin{bmatrix} y_{i,t+1} \\ x_{i,t+1} \end{bmatrix} = \begin{bmatrix} \mu_i \\ 0 \end{bmatrix} \begin{bmatrix} a_t & b_t \\ 0 & \rho_t \end{bmatrix} \begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix} + \begin{bmatrix} u^y_{i,t+1} \\ u^x_{i,t+1} \end{bmatrix}, \tag{49}$$

Where the error terms satisfy:

$$\begin{bmatrix} u^y_{i,t+1} \\ u^x_{i,t+1} \end{bmatrix} \overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & 1 \end{bmatrix} \right). \tag{50}$$

In this equation and all further equations in this section, the μ_i is the fixed effect of each unit i. a_t is the parameter corresponding with the lagged dependent variable. b_t is the paramater corresponding with the explanatory variable. In experiment 1 DGP 1 and 2 have discrete time breaks in the parameters a_t and b_t respectively. In DGP 3 and 4 a_t and b_t are multiplied by $\frac{t}{T}$. In DGP 5 and 6 they are multiplied by $\frac{t}{T}^2$ and lastly, in DGP 7 and 8 a_t and b_T follow a random walk. In all DGP's the fixed effects β_i are equal to i, thus $\beta_1 = 1$ and $\beta_2 = 2$ etc. Whereas Inoue Panel is the method of Inoue et al. (2016) applied on a panel data model, Inoue Single employs the method of Inoue et al. (2016) performed on single time series. Furthermore, Full Sample Panel and Full Sample Single are a panel data model and single time series model estimated using the full sample. A total of 100 simulations are done, with j denoting which simulation. The RMSFE's given in the table are calculated as the average of the RMSFE's over all simulations and units: $RMSFE = \sum_{j=1}^{100} \sum_{i=1}^{N} RMSFE_{i,j}/100N$ and the RMSFE is calculated as follows: $\sqrt{(y_{T+1} - \hat{y_{T+1}})^2}$ with $\hat{y_{T+1}}$ the estimate using corresponding model and y_{T+1} the actual observation.

Table 1: RMSFE's Experiment 1

DGP	Full Sample Panel	Full Sample Single	Inoue Panel	Inoue Single
1	1.358	1.482	**0.818**	1.206
2	0.975	1.000	**0.840**	0.967
3	1.286	1.233	**0.965**	1.290
4	1.110	1.088	**0.905**	1.113
5	1.410	1.456	**1.117**	1.391
6	1.540	1.484	**0.980**	1.544
7	1.882	1.914	1.900	**1.871**
8	1.185	1.179	**0.984**	1.187

Table 1 shows the RMSFE results of experiment 1. The results confirm the expectation that the usage of the Inoue Panel improves the RMSFE's. Due to the extra information available by combining multiple individuals, the estimates of the parameters will be more accurate which results in better forecasting. Except for DGP 7, where *at* follows a random walk, here the Full Sample Panel performs best. This might be explained due to the fact that the Inoue Panel method works best when the change in the parameters is linear, or at least when there is a trend. In a random walk a trend is obviously not present and coincidental trends in the random walk might set the method on the wrong foot. Another noteworthy aspect is the moving window length that the Inoue Panel and Inoue Single methods choose.

Table 2: Average Moving Window Experiment 1

DGP	R Single	R Panel
1	**18**	19
2	**19**	20
3	16	**7**
4	16	**10**
5	15	**6**
6	15	**8**
7	**17**	18
8	17	**16**

In this table, the average moving window that is chosen using the Inoue Panel and Inoue Single methods are displayed, with results rounded off to the nearest integer. Interestingly to see and echoing the essence of this paper, is the observation that the Inoue Panel method includes significantly less time observations in the estimation of the model, due to the fact that it has more units and therefore more observations even when using a short window to estimate the parameters with.

As the results of experiment 1 are as anticipated, in experiment 2, the method is tested on a generated data set that is potentially closer to reality. Using the same DGP with a slight adjustment yields:

$$\begin{bmatrix} y_{i,t+1} \\ x_{i,t+1} \end{bmatrix} = \begin{bmatrix} \mu_i \\ 0 \end{bmatrix} + \begin{bmatrix} a_{it} & b_{it} \\ 0 & \rho_t \end{bmatrix} \begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix} + \begin{bmatrix} u^y_{i,t+1} \\ u^x_{i,t+1} \end{bmatrix}, \tag{51}$$

Where the error terms still satisfy:

$$\begin{bmatrix} u^y_{i,t+1} \\ u^x_{i,t+1} \end{bmatrix} \overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & 1 \end{bmatrix} \right). \tag{52}$$

Now rather than having equal parameters among units, each unit will have its own parameters a_{it} and b_{it}, which are drawn uniformly between 0.1 and 0.9 for a_{it} and 0 and 2 for b_{it} (the choice of the exact values here is arbitrary). Using this DGP 9 to 14 are created, which are identical to DGP 3-8, yet with the newly defined a_{it} and b_{it}.

Since parameters are generated at random in this experiment, clustering all units together and assuming they are homogeneous is most likely a false assumption. Therefore, the methods to create clusters are first applied to determine which units can be clustered together, and on these resulting clusters the algorithm of Inoue is applied. To create the clusters, I will consider the cluster method based on a fixed amount of observations and the dynamic clustering method. For the method that uses a fixed set of observations it is decided to use 20 observations to calculate the clusters. Then similarly in the dynamic cluster a selection value in the AICa of 20 is used. Both methods iterate over 100 starting values to find clusters, this might not be the optimal amount to find correct clusters as the method is more likely to find a local minimum. Yet, iterating over more than 100 starting values will slow the algorithm down too much. Results shown might thus not even be at the fullest potential.

Table 3: RMSFE's Experiment 2

DGP	Inoue Single	FS Single	Inoue Panel	Inoue Panel DC	FS Panel	FS Panel DC
9	**1.018**	1.086	0.938	**0.910**	**1.091**	1.180
10	**1.060**	1.101	0.978	**0.963**	**1.098**	1.211
11	**1.076**	1.260	1.045	**0.996**	**1.243**	1.297
12	**1.113**	1.471	1.082	**1.061**	**1.445**	1.574
13	1.195	**1.146**	1.135	**1.115**	**1.142**	1.263
14	1.202	**1.155**	1.124	**1.095**	**1.138**	1.207

DC here stands for Dynamic clustering, and FS stands for full sample.

From the results shown in table 3 it can be concluded that the use of panel data models improves the RMSFE across all DGPs. This implies that even in the scenario where the parameters are drawn at random the method will find similar behaving units, cluster them together and create better forecasts. The extra information added by clustering the units outweighs the bias the introduced by wrongly assuming units to be homogeneous. Furthermore, by looking at the middle two columns it can be seen that the dynamic cluster method outperforms the cluster based on fixed time period across all DGPs as well. Remarkably, the clusters based on the dynamic cluster method perform worse when they are used in the full sample. The dynamic cluster method thus seems to make clusters that are more tailored to the time varying properties of the parameters. Thereby voicing the purpose of making dynamic clusters. Moreover, in table 4, it can be seen that the dynamic cluster method results in the use of a smaller moving window in the final estimation of the model. This might be in line with the average amount of clusters that are created, given that the average amount for the static cluster method lies between 5 and 7 over the DGPs compared to 4 and 6 in the case of dynamic cluster. This implies that, on average, the dynamic cluster method will have more units per cluster, thereby allowing it to include less time points in the estimation of the model, which in turn will estimate a parameter closer to the current true parameter and hence yielding improved forecasts. Either way, when using the panel data model, on average, the amount of units per clusters should be 2, which still substantially improves the RMSFE's.

Table 4: Average Moving Window Experiment 2

DGP	R Single	R panel	R panel DC
9	15	14	**13**
10	15	16	**14**
11	15	12	**11**
12	14	13	**12**
13	**16**	17	**16**
14	**15**	17	16

In experiment 1 it was very obvious that the Inoue Panel method included fewer time points in the estimation of the model. However, in experiment 2 such an observation seems to be less clear, although it is still present. On the contrary, in the random walk models such an observation does not seem to be present at all.

In the experiment 2 the amount of observations to base the clusters on and the Selection Value were chosen arbitrarily. Moreover, all the units in each DGP showed the same time varying pattern, which can potentially steer towards a result that is as expected. Therefore, I created another experiment (3), where the DGP for every unit draws the parameters uniformly, and consequently draws the time varying pattern it will follow randomly. The data set created is, therefore, a bunch of randomness with possibly a slight pattern that can be recognized. On this data set, I test both cluster methods for different fixed T and Selection Value values, to see which settings perform the best. The methods are tested over 10 rather than 100 sets, due to the high computational time it takes to run every iteration of the methods. The data set is created with 20 units and 40 time periods.

Table 5: RMSFE's Experiment 3

T \ SV	Inoue Panel	Inoue Panel DC	FS Panel	FS Panel DC
5	1.000	**0.966**	1.321	1.279
10	**0.968**	1.000	1.259	1.290
20	**0.951**	1.030	1.230	1.350
30	1.055	**1.014**	1.257	1.331
40	1.089	**1.015**	1.223	1.317

The method that creates clusters based on a fixed time period of 20 observations seems to outperform other methods. In turn an SV value of 20 is the value of SV that performs worst. Although this result does not turn out favorably for the DC method, it cannot be concluded that there is a significant difference between the two methods in performance; when checking the best and worst performance (T/SV = 20) there is no significant difference in the two methods (T = 0.87).

3.2 Static Panel Data Model

Here I will evaluate the Static Panel data model, which does not include an autoregressive component. Also in this section the data generated contains 10 units and 40 time periods.
The DGP in experiment 4 follows the following process:

$$\begin{bmatrix} y_{i,t+1} \\ x_{i,t+1} \end{bmatrix} = \begin{bmatrix} \mu_i \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & b_t \\ 0 & \rho_t \end{bmatrix} \begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix} + \begin{bmatrix} u^y_{i,t+1} \\ u^x_{i,t+1} \end{bmatrix}, \tag{53}$$

Where the error terms satisfy:

$$\begin{bmatrix} u^y_{i,t+1} \\ u^x_{i,t+1} \end{bmatrix} \overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & 1 \end{bmatrix} \right). \tag{54}$$

In the static simulation, the dependent variable will be explained by only one independent variable. This simplifies the model substantially and, furthermore, avoids a requirement for any bias correction for the estimates. Since only one variable is included, 4 DGPs will be considered: (1) a discrete break, (2) varying over $\frac{t}{T}$, (3) varying over $\frac{t}{T}^2$, and (4) a random walk.

Table 6: RMSFE's Experiment 4

DGP	Inoue Panel	Inoue Single	Full Sample Panel	Full Sample Single
15	**0.5713**	0.6328	0.9524	0.7439
16	0.8850	**0.8656**	0.9195	1.0184
17	**1.2426**	1.3884	1.9829	1.8205
18	**0.9274**	1.0709	1.6611	1.6268

Table 6 shows that the using the Inoue Panel method on the Full Sample Panel model gives better results compared to the Inoue Single method. Yet the improvement in RMSFE is not as substantial when compared to the dynamic model. The result that the Inoue Single method works better for DGP 16 is unexpected, since the method included less data points for the estimation of the parameters by which the estimates should be less certain and consequently provide less accurate forecasts. Yet the opposite seems to hold in the simulated data set.

Similar to the dynamic model, also in the Static Panel data model a more realistic setting is created by drawing the parameter of the model uniformly between 0 and 2, consequently drawing the time variance from the DGPs with the parameter varying by $\frac{t}{T}$, $\frac{t}{T}^2$ and the parameter being a random walk, creating experiment 5. Similar to the dynamic model the clusters are estimated using the AIC and AICa information criteria over 100 different starting values, using 20 observations for the cluster method with fixed observations and an SV value of 20 for the dynamic cluster.

Table 7: RMSFE's Experiment 5

DGP	Inoue Single	FS Single	Inoue Panel	Inoue Panel DG	FS Panel	FS Panel DG
19	0.938	1.141	**0.911**	0.933	1.120	1.199
20	0.936	1.582	0.943	**0.924**	1.566	1.650
21	0.993	1.162	0.986	**0.981**	1.126	1.257

Table 7 shows, once more, that the Inoue Panel method outperforms the other methods, although the results are not as convincing as the dynamic model. This might be due to the incorporation of only one variable, which makes estimates on this variable converge to the true value quickly, meaning that extra information will not necessarily improve results. However, in this experiment it is expected that the method includes less data points as the model is less complex. This does not seem to be the case. The average moving windows for DGP 15 to 21 are found in table 8.

Table 8: Optimal moving windows experiment 4&5

DGP	R Panel	R Single
15	**17**	20
16	**12**	15
17	**7**	8
18	**7**	11
19	**12**	14
20	**10**	12
21	**14**	16

To conclude, improvement in almost every DGP proposed in this section is observed. Improvements in the dynamic section were more apparant compared to the static section. However, the combination of methods thus far performs as anticipated.

4 Empirical Experiment

To further explore the combination of methods, an experimental study on an empirical data set, consisting of 16 OECD countries' GDP growth data, is conducted. Per country eight variables are retrieved: (1) GDP growth, (2) short term interest, (3) long term interest, (4) industrial production, (5) consumer price index, (6) production price index, (7) M1 and (8) unemployment rate. The data runs from the first quarter of 2002 to the last quarter of 2018 on a quarterly basis. The 16 included countries are European Union countries that had a complete data set for the selected variables over the specific time period. Ireland was, however, left out from the data, since it contained unpredictable outliers in the data due to inflated GDP growth from investments for tax reasons, this can be seen in table 17 where the variance of GDP growth is very high. This leaves 15 use-able countries. There are a total of 64 time observations per country. Each time series is reduced by one observation, since the model performs one step ahead forecasts, leaving the first 63 independent variables to estimate the last 63 dependent variables. Moreover, similarly as in Inoue et al. (2016) for the variables industrial production, consumer price index, production price index, M1 and unemployment rate log first differences are taken as variables in the model, which demand two observations to be created, reducing the total time observations in the model to 62. This means that the time series model of each country will run from third quarter of 2002 to the last quarter of 2018, based on independent variables of one quarter prior. If each country would have its own time series, including intercept and dynamic component, then, to estimate the model, a total of 9 parameters would have to be estimated per country. Considering the method of Inoue et al. (2016) to select the optimal number of observations, estimating these 9 parameters seems impossible when using too few observations, which hence, reveals the potential benefit of using panel data modelling.

Before determining the optimal moving window, countries have to be clustered. In this empirical section the countries will be clustered in three different manners: Firstly, the clusters are based on a fixed amount of 20 observations. Secondly, the clusters are based on the Dynamic cluster method as described in the methodology. A selection value (SV) of 20 is chosen for the dynamic cluster method. Lastly, clusters will also be made using the full sample of data. In the empirical section the optimal moving window is determined over the interval of 10 to 40 observations. Since the model includes 9 variables the minimum of 10 observations to estimate the model seems reasonable. To test the performances, the models will run iteratively over the last 20 observations performing one step ahead forecasts.

In short, the following steps are taken: the model will create one step ahead forecasts over 20 time periods by first creating a cluster of the units, and then per cluster performing the optimal window selection method. Results are shown in table 9.

Table 9: RMSFE's Empirical data

	mean	variance
Full Panel	0.469	0.412
Full Single	0.525	0.440
Inoue Panel	0.523	0.531
Inoue Single	1.100	3.937
Full Panel DC	0.489	0.262
Inoue Panel DC	0.468	0.204
Full Panel FPC	0.467	0.257
Inoue Panel FPC	0.501	0.283

FPC in the table stands for Full Panel Clustering

The first observation that is noticeable from the results concerns the weak performance of the Inoue Single method; it performs by far the worst, about twice as bad as the Full Single method. These results do not support the Inoue method. Secondly, it seems that the Dynamic Clustering method outperforms all other methods, expressed in the lowest variance in the RMSFE and the lowest on average RMSFE mean between Inoue and Full Panel method. Out of the results, only the Inoue Single method is significantly worse compared to the rest, however, the other results do not differ significantly from one another.

As can be seen in the results, the results using panel data improve performance compared to the single time series forecasting, albeit slightly when compared to the Full Single performance. A make or break issue when using panel data is finding correct clusters, where the three cluster methods, using 20 observations, dynamic clustering and using the full sample use 5.2, 2.6 and 5 cluster on average respectively. It is hard to pinpoint the performance of different amount of clusters on the RMSFE, however, the Dynamic Clustering method uses noticeably less clusters in the method. This can be due to the extra bias that is most likely present in the AICa measure, which puts more relative weight on the total parameters in the model, steering it to include less clusters to have less parameters. The fact that this Dynamic Clustering method performs the best overall, with lowest variance and lowest average RMSFE among Inoue and Full Panel models, suggests that penalizing extra parameters in the model might improve performance, thus using the BIC measure rather than the AIC will reduce the amount of clusters and might prove beneficial in this specific empirical experiment.

Table 10: RMSFE's using BIC clustering

	mean	variance
Full Panel	0.467	0.217
Inoue Panel	0.446	0.188
Full Panel DC	0.491	0.249
Inoue Panel DC	0.427	0.165
Full Panel FPC	0.462	0.161
Inoue Panel FPC	0.463	0.207

Table 10 shows the performance of the methods when the BIC measure is used in selecting the number of clusters to be used. It can easily be observed that almost all methods improve in performance when the BIC is used. However, once again, this cannot be significantly substantiated. The number of groups included when using the BIC measure are 1.7, 1.95 and 2 for the 20 observation, dynamic clustering and full sample clustering methods respectively. The 20 observation method will never use more than 4 clusters, the dynamic clustering method will never use more than 3 clusters and both the methods will occasionally put all units in one cluster. The full sample clustering method will always divide the units into two clusters. These results seem to suggest that assuming all units to be in the same cluster will provide the best results, although this is not the case with a RMSFE of 0.738 when the full sample is used to estimate the model and a RMSFE of 0.683 when the Inoue method is used. This indicates that the cluster methods actually improve performance.

Table 11: Average window length per method

	AIC	BIC
Inoue Single	22.48	
Dynamic clustering	14.73	14.89
20 observation clustering	19.20	17.25
full sample clustering	20.42	17.70

Moving on to the amount of time observations included in the estimation of the models. Table 11 shows the average window lengths used. If these results are related back to the observed RMSFE's it is evident that the dynamic clustering method includes the least time observations, where the dynamic clustering method also provides the best performing RMSFE's. This in line with the hypothesis that by including less time observations, less outdated data is included in estimating the parameters, resulting in less biased estimated, which results in better forecasts. The usage of Panel Data really shows here, as the added observations from multiple units makes it possible to reduce the time dimension while keeping the estimates consistent.

5 Discussion

The aim of this thesis was to find and test a method to combine the method of optimal window selection of Inoue et al. (2016) with panel data. By making use of a dynamic fixed effects model the optimal moving window method can easily be applied on panel data. A potential problem concerned the bias of estimating a panel data model introduced when the time dimension of the data set is low. Even though a method to cope with this problem was found, the method turned out to be computationally demanding. Moreover, in the empirical section of this thesis, the bias correction does not provide any necessary addition and was therefore left out. What did add value to this thesis by making the results more widely applicable in empirical scenarios, was the creation of clusters in the panel data, which circumvented unnecessarily assuming different units to be homogeneous when they are not. From the three methods included in this thesis, the dynamic clustering method appeared to perform the best in the empirical application, when combined with the BIC information criterion to find the optimal number of groups to include in the estimation of the model. Although the improvement compared to simpler methods was slight, the improvement nonetheless supports the hypothesis that the usage of panel data is valuable when dealing with time varying parameters. This result was especially apparent in the Monte Carlo simulation, as the newly proposed method outperformed the alternatives in almost every scenario.

An important finding of this thesis is that the usage of the method of Inoue et al. (2016) did not provide any improvements when applied on single time series. In fact, the method performed significantly worse compared to any other method used in this thesis.

This finding, simultaneously poses one of the biggest limitations of this study. As the method of Inoue et al. (2016) cannot improve model performance on single time series, improvements regarding the application on panel data must be limited. Apparently, the time varying property of the time series was not captured well in this particular data set. However, in case future research uses a data set that follows the underlying assumptions of time variance more strictly, the combination of methods can potentially provide significant improvements on primitive methods.

Another limitation of this study is the computational time it takes to perform the combination of methods. This is especially apparent when trying to find clusters in the panel data. The method to find optimal moving window iterates over many potential moving windows when estimating the parameters to finally arrive at an optimal window. When performing the dynamic clustering method, this iteration over potential windows has to be done for every combination of units and every number of clusters, which makes the method iterate over a method that iterates over estimating the parameters numerous times. This results in a very slow performing algorithm, which for the empirical part of this studies took over 90 minutes to run. In case the data set is expanded to include more units and more time observations, the method will run for even longer. This makes the method not very practical to use for individuals with limited computing power.

Potential improvements in this regard can be made, by finding methods that find optimal moving windows, which are less computationally demanding. Another area susceptible to improvement concerns the cluster selection method. Moreover, if the researcher predetermines the number of clusters to include in the estimation the method does not have to iterate over all potential number of clusters, which relieves it from a computational burden even more.

In my thesis I decided that it was not useful to include the bias correction method, since it did not provide useful improvements in performance. However, in potential scenarios where it might prove useful, the bias correction method proposed in this thesis is not practical. If clusters are to be found when estimates need to be corrected with the bootstrap bias method, the method will iterate endlessly. Therefore, future research might focus on a more computationally friendly bias correction method. However, it is important to assess whether including a bias correction is even purposeful.

To conclude, when optimal window selection methods are used on panel data models with time varying parameters, it gives more room to lower the time dimension, since the estimates will stay consistent due to the many units included in the estimation of the parameters. This fact could potentially steer the model to include less time observations when forecasting. The parameters for forecasting are then based on more recent observations, which are closer to the current 'true' parameters and will thus improve forecasts. Drawbacks include both wrongly assuming units to be homogeneous and the potential bias in the estimates when the time dimension is low. Both can be adjusted for by applying a clustering method and a bias correction method. I have shown that clustering methods that take into account the time varying property of the parameters outperform other methods and that when the time dimension is not too small a bias correction will not be necessary.

Appendix A

Table 12: Data Generating Processes

DGP	μ_i	σ	a_t	b_t	comments
1	1 to 10	1	$0.9 - 0.4I(t > 0.5T + 1)$	1	discrete break
2	1 to 10	1	0.9	$1 + I(t > 0.5T + 1)$	discrete break
3	1 to 10	1	$0.9 - 0.4\frac{t}{T}$	1	linearly time varying
4	1 to 10	1	0.9	$1 + \frac{t}{T}$	linearly time varying
5	1 to 10	1	$0.9 - 0.4\frac{t}{T}^2$	1	quadratically time varying
6	1 to 10	1	0.9	$1 + \frac{t}{T}^2$	quadratically time varying
7	1 to 10	1	$a_t = a_{t-1} + \frac{0.1}{\sqrt{T}}\epsilon_t$	1	random walk in a_t
8	1 to 10	1	0.9	$b_t = b_{t-1} + \frac{1}{\sqrt{T}}\epsilon_t$	random walk in b_t
9	1 to 10	1	$a_{it} - 0.4\frac{t}{T}$	b_{it}	$a_{it} \sim unif(0.1, 0.9), b_{it} \sim unif(0, 2)$
10	1 to 10	1	a_{it}	$b_{it} + \frac{t}{T}$	$a_{it} \sim unif(0.1, 0.9), b_{it} \sim unif(0, 2)$
11	1 to 10	1	$a_{it} - 0.4\frac{t}{T}^2$	b_{it}	$a_{it} \sim unif(0.1, 0.9), b_{it} \sim unif(0, 2)$
12	1 to 10	1	a_{it}	$b_{it} + \frac{t}{T}^2$	$a_{it} \sim unif(0.1, 0.9), b_{it} \sim unif(0, 2)$
13	1 to 10	1	$a_{it} = a_{i,t-1} + \frac{0.1}{\sqrt{T}}\epsilon_t$	b_{it}	$a_{it} \sim unif(0.1, 0.9), b_{it} \sim unif(0, 2)$
14	1 to 10	1	a_{it}	$b_{it} = b_{i,t-1} + \frac{0.1}{\sqrt{T}}\epsilon_t$	$a_{it} \sim unif(0.1, 0.9), b_{it} \sim unif(0, 2)$
15	1 to 10	1	0	$1 + I(t > 0.5T + 1)$	discrete break
16	1 to 10	1	0	$1 + \frac{t}{T}$	linearly time varying
17	1 to 10	1	0	$1 + \frac{t}{T}^2$	quadratically time varying
18	1 to 10	1	0	$b_t = b_{t-1} + \frac{1}{\sqrt{T}}\epsilon_t$	random walk in bt
19	1 to 10	1	0	$b_{it} - 0.4I(t > 0.5T + 1)$	$b_{it} \sim unif(0, 2)$
20	1 to 10	1	0	$b_{it} - 0.4\frac{t}{T}$	$b_{it} \sim unif(0, 2)$
21	1 to 10	1	0	$b_{it} - 0.4\frac{t}{T}^2$	$b_{it} \sim unif(0, 2)$

Table 13: Estimates with and without bootstrap bias correction

N		2		5		10		20	
T		bias bt	bias at						
5	LSDV	**-0.007**	-0.178	-0.015	-0.067	-0.023	-0.039	**0.001**	-0.011
	LSDVc	0.014	**-0.014**	**-0.007**	**0.024**	**-0.020**	**0.007**	0.004	**0.005**
10		**0.006**	-0.035	0.015	-0.027	0.003	-0.013	**0.000**	-0.005
		-0.008	**0.013**	**0.003**	**0.003**	**-0.001**	**0.002**	-0.002	**0.001**
15		**0.007**	-0.026	0.021	-0.017	0.008	-0.010	0.005	-0.004
		-0.019	**0.001**	**0.009**	**-0.002**	**0.001**	**-0.002**	**0.003**	**-0.001**
20		0.019	-0.014	0.007	-0.010	**0.003**	-0.005	0.002	-0.002
		0.002	**0.004**	**-0.004**	**0.000**	-0.004	**0.000**	**0.000**	**0.000**
25		0.008	-0.010	0.017	-0.008	0.013	-0.004	0.008	-0.001
		-0.008	**0.001**	**0.007**	**0.000**	**0.007**	**0.000**	**0.006**	**0.001**
30		0.022	-0.013	0.010	-0.006	0.010	-0.003	0.007	-0.002
		0.008	**-0.003**	**0.001**	**-0.001**	**0.004**	**0.000**	**0.005**	**0.000**

Table 14: Experiment 2 RMSFE's clustered panel models

Inoue	AIC	AICa	BIC	BICa	HQE	HQEa
1	0.940	0.916	0.937	0.919	0.948	**0.912**
2	0.977	**0.951**	0.963	0.960	0.962	0.957
3	1.043	0.982	1.031	0.988	1.030	**0.980**
4	1.083	1.052	1.033	**1.051**	1.058	1.057
5	1.135	**1.099**	1.113	1.116	1.120	1.106
6	1.125	**1.115**	1.134	1.117	1.129	1.122
Full Sample						
1	**1.095**	1.179	1.134	1.208	1.121	1.199
2	**1.100**	1.212	1.127	1.227	1.114	1.213
3	1.248	1.303	**1.244**	1.327	1.247	1.313
4	**1.446**	1.561	1.475	1.583	1.453	1.584
5	**1.143**	1.263	1.155	1.268	1.145	1.269
6	**1.138**	1.231	1.167	1.239	1.146	1.231

Appendix B

Table 15: Variable Summaries 1

	aut		bel		cze		deu		dnk		esp	
	mean	variance										
1	0.387	0.423	0.367	0.289	0.712	0.875	0.343	0.776	0.284	0.903	0.356	0.477
2	1.613	2.446	1.613	2.446	1.670	1.385	1.613	2.446	1.822	2.807	1.613	2.446
3	2.949	2.063	3.137	1.980	3.177	2.406	2.618	2.250	2.790	2.320	3.796	1.678
4	0.007	0.00033	0.006	0.00052	0.008	0.00048	0.005	0.00066	-0.001	0.00064	-0.002	0.00038
5	0.005	0.00002	0.005	0.00002	0.005	0.00006	0.004	0.00001	0.004	0.00002	0.005	0.00012
6	0.004	0.00008	0.006	0.00045	0.002	0.00016	0.003	0.00006	0.005	0.00007	0.005	0.00021
7	0.020	0.00014	0.020	0.00014	0.026	0.00011	0.020	0.00014	0.016	0.00029	0.020	0.00014
8	0.021	0.095	-0.004	0.192	-0.076	0.107	-0.074	0.068	0.024	0.128	0.092	0.502

1 = GDP growth, 2 = short term interest, 3 = long term interest, 4 = industrial production, 5 = consumer price index, 6 = production price index, 7 = M1, 8 = unemployment rate

Variable 4 to 8 are log differences, therefore negative mean at variable 8 does not mean negative unemployment rate.

Table 16: Variable Summaries 2

	fin		grc		irl		ita		lva	
	mean	variance								
1	0.312	1.638	-0.129	2.645	1.211	13.265	0.029	0.496	0.824	3.996
2	1.613	2.446	1.613	2.446	1.613	2.446	1.613	2.446	3.372	15.086
3	2.845	2.094	7.837	26.899	4.075	5.025	3.894	1.483	4.810	9.871
4	0.001	0.00076	-0.004	0.00053	0.010	0.00334	-0.002	0.00051	0.009	0.00083
5	0.003	0.00002	0.004	0.00020	0.004	0.00008	0.004	0.00001	0.009	0.00019
6	0.004	0.00020	0.005	0.00081	0.003	0.00019	0.004	0.00015	0.009	0.00027
7	0.020	0.00014	0.020	0.00014	0.020	0.00014	0.020	0.00014	0.020	0.00014
8	-0.006	0.086	0.166	0.491	0.037	0.323	0.041	0.118	-0.071	1.305

1 = GDP growth, 2 = short term interest, 3 = long term interest, 4 = industrial production, 5 = consumer price index, 6 = production price index, 7 = M1, 8 = unemployment rate

Table 17: Variable Summaries 3

	nld		nor		pol		prt		swe	
	mean	variance								
1	0.343	0.500	0.391	1.121	0.976	0.889	0.093	0.579	0.542	0.986
2	1.613	2.446	2.867	3.004	4.293	3.208	1.613	2.446	1.512	2.295
3	2.832	2.071	3.435	1.862	5.076	1.820	4.961	5.825	2.841	2.130
4	0.002	0.00052	-0.003	0.00086	0.013	0.00039	-0.003	0.00045	0.001	0.00059
5	0.004	0.00003	0.005	0.00005	0.005	0.00005	0.004	0.00007	0.003	0.00003
6	0.005	0.00048	0.007	0.00010	0.005	0.00020	0.003	0.00022	0.004	0.00008
7	0.020	0.00014	0.026	0.00492	0.032	0.00032	0.020	0.00014	0.019	0.00011
8	0.036	0.053	0.004	0.038	-0.239	0.227	0.071	0.230	0.028	0.107

1 = GDP growth, 2 = short term interest, 3 = long term interest, 4 = industrial production, 5 = consumer price index, 6 = production price index, 7 = M1, 8 = unemployment rate

Table 18: RMSFE results HQI

	mean	var	R	#groups
Full Panel HQI	0.467	0.387		3.35
Inoue Panel HQI	0.512	0.500	17.67	"
Full Panel HQI DC	0.494	0.297		2.25
Inoue Panel HQI DC	0.473	0.243	14.82	"
Full Panel HQI FSC	0.467	0.248		4.65
Inoue Panel HQI FSC	0.497	0.282	20.49	"

References

Anderson, T., & Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, *18*, 47-82.

Baltagi, B. H. (2012). *Econometric analysis of panel data.*

Bester, C. A., & Hansen, C. B. (2009). Grouped effects estimators in fixed effects models. *Journal of Econometrics*.

Bonhomme, S., & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*.

Bruno, G. S. F. (2005). Estimation and inference in dynamic unbalanced panel-data models with a small number of individuals. *The Stata Journal*, *5*, 473-500.

Bun, M. J., & Carree, M. A. (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business Economics Statistics*, *23:2*, 200-210.

Bun, M. J., & Kiviet, J. F. (2003). On the diminishing returns of higher-order terms in asymptotic expansions of bias. *Economics Letters*, *79*, 145-152.

Clements, M. P., & Hendry, D. F. (1998). Forecasting economic processes. *International Journal of Forecasting*.

Everaert, G., & Pozzi, L. (2007). Bootstrap-based bias correction for dynamic panels. *Journal of Economic Dynamics Control*, *31*, 1160-1184.

Inoue, A., Jin, L., & Rossi, B. (2016). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*.

Judson, R. A., & Owen, A. L. (1999). Estimating dynamic panel data models: A guide for macroeconomists. *Economics Letters*, *65*.

Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics*, *68*, 53-78.

Lin, C. C., & Ng, S. (2012). Estimation of panel data models with parameter heterogeneity when group membership is unknown. *Journal of Econometric Methods*.

Lu, H., & Huang, S. (2011). Clustering panel data. *Data Mining for Marketing*, *1*.

Manuel Arellano, S. B. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*, 277-297.

Mark, N. C., & Sul, D. (2011). When are pooled panel-data regression forecasts of exchange rates more accurate than the time-series regression forecasts? *Handbook of exchange rates*.

M. Hashem Pesarana, A. T. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, *137*, 134-161.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, *16*.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, *49*.

Paye, B. S., & Timmermann, A. (2005). Instability of return prediction models. *Journal of Empirical Finance*.

Pesaran, M. H., Pick, A., & Pranovich, M. (2011). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, *177*.

Pesaran, M. H., & Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, *137*.

Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature*.

Stock, J., & Watson, M. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*.

Stock, J. H., & Watson, M. W. (2007). Why has u.s. inflation become harder to forecast? *Journal of Money, Credit and Banking*, *39*.