Empirical Methods for Causal Inference and Forecasting in Panel Data Settings

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2026

vorgelegt von
Lennart Bolwin, M.Sc.,
aus Mainz

Referent: Prof. Dr. Jörg Breitung
Koreferent: Prof. Dr. Tom Zimmermann
Tag der Promotion: 22.07.2026

Vorwort

Mein größter Dank gilt meinem Betreuer, Professor Dr. Jörg Breitung, ohne den diese Dissertation nicht möglich gewesen wäre. Besonders hervorheben möchte ich, dass ich in ihm einen Betreuer hatte, der mich seit Beginn meines Masterstudiums in meinem Interesse für die Ökonometrie gefördert und mir durch sein Streben nach eleganten statistischen Lösungen sowie seinen großen Erfahrungsschatz zahlreiche wertvolle Impulse vermittelt hat. Ebenfalls bedanke ich mich bei meinen Koautoren und meinem Zweitbetreuer, Professor Dr. Tom Zimmermann.

Zuletzt gilt mein besonderer Dank meinen Freunden und meiner Familie, die mir in vielen Gesprächen immer ein offenes Ohr geschenkt und mich über den gesamten Zeitraum der Promotion unterstützt haben.

Köln, am 14.05.2026
Lennart Bolwin

Table of Contents

List of Figures

List of Tables

1 Introduction

"Trying to predict the future is like trying to drive down a country road at night with no lights while looking out the back window" (usually attributed to Peter Drucker).

Although management scholar Peter Drucker's remark may appear amusing, it captures a central challenge of forecasting. Future outcomes are unknown, and forecasting relies on learning systematic patterns from historical data. In economics, forecasting plays a particularly prominent role because economic quantities are measured using stable concepts over time and directly inform the decision making of individuals, organizations, and governments. This close link between forecasts and economic decisions has made forecasting a central task of empirical economic research. At the same time, economic forecasting poses substantial challenges from a statistical perspective. These challenges arise not only from limited data availability and structural instability, but also from the need to balance model flexibility, robustness, and interpretability.

A central concern in statistical forecasting is external validity. Forecasting models are only informative if they generalize well to out-of-sample data. This naturally introduces uncertainty, both within a given model and across competing model specifications. Within-model uncertainty is typically summarized by forecast intervals, which quantify uncertainty conditional on the model being correctly specified. However, such intervals do not address uncertainty regarding the choice of the model itself. Estimating alternative forecasting models and comparing their out-of-sample predictions is therefore an essential step in assessing model uncertainty.

Uncertainty surrounding the estimated model is particularly important in macroeconomics, where the available data are often characterized by a limited time dimension, which restricts the degrees of freedom available for estimation. As a result, forecasting frequently involves settings in which classical unregularized methods can perform poorly. Dimensionality reduction techniques, such as principal components, and explicit regularization strategies are therefore commonly employed to stabilize estimation and improve predictive performance. In contrast to causal inference, where unbiasedness is often a primary objective, forecasting problems typically benefit from trading a small amount of bias for a larger reduction in

variance. This bias–variance trade-off is central to obtaining robust forecasts and naturally links forecasting to more data-driven approaches commonly associated with machine learning. At the same time, interpretability is frequently required in economic forecasting, calling for regularized methods that combine predictive accuracy with transparent and economically meaningful model structures.

A different perspective arises when repeated observations of the same units are available over time, as in panel data settings. Despite their prevalence in applied work, the literature on forecasting with panel data remains relatively limited. A naïve approach treats each cross-sectional unit independently, thereby ignoring valuable information contained in the panel structure. From a statistical viewpoint, more efficient approaches exploit both the cross-sectional and time-series dimension jointly. In dynamic panel models, parameters such as autoregressive and slope coefficients can, under suitable conditions, be estimated precisely using information from both dimensions. This shifts attention toward individual heterogeneity, particularly unit-specific intercepts, which enter the model in every period and thus substantially affect forecast accuracy. Accurately accounting for such heterogeneity is therefore crucial for obtaining reliable forecasts in panel data environments.

The rest of this thesis is structured as follows. Chapter 2 is joint work with Jörg Breitung and Justus Töns and contributes to the literature on synthetic control methods, which have become important tools for policy evaluation in recent years. At its core, the synthetic control method can be interpreted as a forecasting problem, as the objective is to construct a counterfactual trajectory that robustly predicts the outcome of the treated unit in the absence of the intervention. The chapter builds on this framework and extends it along several dimensions that are particularly relevant for empirical applications with limited data and dynamic time-series features. The central contribution of the chapter is the development of a regularized synthetic control estimator that combines the core ideas of the original approach with Ridge-type shrinkage. Individual donor weights are shrunk toward zero while their sum is simultaneously shrunk toward one, preserving the economic interpretability of the original synthetic control estimator while yielding a closed-form solution that is fast to compute and tune. An important advantage of the proposed estimator is its Bayesian representation, which allows for a coherent assessment of estimation uncertainty through credibility intervals. This is particularly important when treatment effects are evaluated over time to estimate the joint effect. Beyond the static setting, the chapter extends the synthetic control framework to dynamic environments by augmenting the donor pool with lagged values of both the donor units and the treated series. This extension enables the method to accommodate serial dependence, non-stationary time series, and cointegrated relationships, a case where the original synthetic control estimator is inconsistent. Monte Carlo simulations and an

empirical application illustrate the performance of the proposed approach and its relevance for applied policy evaluation. The core idea of the REGSC estimator was developed by Jörg Breitung, who also contributed substantially to the theoretical framework and the design of the Monte Carlo study. Justus Töns contributed to the development of the dynamic extension and the corresponding simulation design. All remaining parts of the analysis, including further methodological development, implementation, empirical application, simulations, and writing, were carried out by the author.

Chapter 3 is joint work with Rouven Haschka and studies macroeconomic forecasting under model uncertainty using dynamic model averaging (DMA). The chapter is motivated by the observation that professional GDP forecasts remain a central input for economic decision making, while the information set available for forecasting has expanded substantially through high-frequency data sources such as online search activity. This raises the question of how such information can be incorporated flexibly. The chapter adopts a twofold forecasting perspective: first, German GDP growth is forecasted within a DMA framework; second, the DMA structure is used as a diagnostic tool to identify variables that contain predictive information but are not fully reflected by experts. DMA is based on a state-space representation estimated via the Kalman filter, where both regression coefficients and model weights are allowed to vary over time. Forgetting factors govern the trade-off between past and current information, enabling the model to adapt to structural changes. Building on this framework, the chapter introduces a parsimonious extension that allows for regressor-specific coefficient variability. This modification enables DMA to better capture heterogeneous dynamics across macroeconomic predictors and improves its stability and performance when combining variables with different sampling frequencies and persistence patterns. A central contribution of the chapter is the first systematic simulation study assessing the forecasting benefits and risks of integrating Google Trends variables as an example of high-frequency online data into DMA. The simulations analyze controlled data-generating processes to isolate how Google Trends affect model averaging behavior across different signal-to-noise regimes, sample sizes, and levels of model complexity. The idea for the methodological extension and the simulation study was developed by the author. Rouven Haschka contributed in particular to the section identifying variables overlooked in expert forecasts and provided valuable guidance, feedback, and revisions throughout the writing process. All remaining parts of the analysis, including model development, implementation, simulations, empirical analysis, and drafting of the manuscript, were carried out by the author.

Chapter 4 is single-authored. I am grateful to Jörg Breitung for very helpful discussions and guidance. The chapter studies forecasting in dynamic panel data models with a large cross-sectional dimension N and a short time dimension T. Such environments are common in

applied work but pose particular challenges for forecasting, as limited time-series information restricts the precise estimation of unit-specific dynamics while cross-sectional heterogeneity plays a central role. The chapter focuses on settings in which homogeneous coefficients can be estimated with high precision, shifting attention toward the estimation of individual heterogeneity and its role for forecast accuracy. The analysis builds on the empirical Bayes framework proposed by Liu et al. (2020), which provides a principled approach to forecasting individual outcomes in dynamic panels by shrinking unit-specific effects toward a common distribution. A key contribution of the chapter is the distinction between stationary and non-stationary initial conditions. Under stationary initial conditions, the underlying process is assumed to have been running long enough that the initial observations are drawn from the stationary distribution. Under non-stationary initial conditions, by contrast, the first observed values are still affected by the starting value. The chapter shows that this distinction is crucial for forecast construction and develops fixed- and random-effects predictors which are subsequently compared to the simple mean and the Tweedie-type forecast of Liu et al. (2020). A comprehensive simulation study confirms that the random-effects predictor under non-stationary initial conditions provides the most reliable forecasting performance across a wide range of scenarios. The design-based Monte Carlo application using German firm-level productivity data confirms the simulation findings and illustrates how additional explanatory variables can be incorporated using a Mundlak-type specification.

This thesis contributes to the econometric literature in several ways. It develops forecasting methods that are computationally fast to estimate, but perform strongly across a wide range of empirical settings. The proposed methods address challenges that frequently arise in applied economic research, including limited data availability, model uncertainty, and heterogeneity. The thesis advances the econometric toolkit for economic forecasting by linking forecasting and policy evaluation through a common perspective. Methods such as synthetic control are treated explicitly as forecasting problems, allowing uncertainty and performance to be assessed in a unified framework. Simulation studies calibrated to realistic data-generating processes provide systematic evidence on the reliability of the proposed approaches. Finally, the methods are applied to economically relevant data sets, including policy evaluation, macroeconomic forecasting, and firm-level panel data. These applications illustrate the usefulness of the proposed methods and offer practical guidance.

2 Alternative Approaches for Estimation and Inference in Synthetic Control Designs

Abstract. The synthetic control (SC) method estimates causal treatment effects by constructing a counterfactual for the treated unit as a weighted average of donor units in the pre-treatment period. To reduce overfitting and support interpretability, the original approach of Abadie, Diamond, and Hainmueller (ADH) restricts weights to be non-negative and sum to one. Building on Doudchenko and Imbens (2016), we introduce a regularized synthetic control estimator (REGSC) that shrinks individual coefficients toward zero and their sum toward one. This combines SC with Ridge-type regularization and yields a closed-form estimator with tunable hyperparameters. The method also has a straightforward Bayesian formulation, enabling credibility intervals for uncertainty quantification. By augmenting the donor pool with lagged values of the donors and the counterfactual, REGSC extends naturally to dynamic settings and accommodates non-stationary and cointegrated series, a case in which the original SC method is inconsistent. Monte Carlo simulations and empirical applications demonstrate that the (dynamic) REGSC estimator outperforms existing SC methods.

2.1 Introduction

Synthetic control methods are purely data-driven and rely on time series and cross-sectional information, forming a counterfactual as a weighted average of untreated units that best matches the treated unit prior to the intervention. In contrast, Pesaran and Smith (2014) adopt a structural time-series perspective, generating counterfactuals from a macroeconomic model estimated solely on pre-intervention data. While synthetic control avoids strong modeling assumptions, the Pesaran–Smith framework is designed for settings without suitable control units and explicitly addresses the Lucas critique.

The synthetic control (henceforth: SC) method was developed by Alberto Abadie, Alexis Diamond, and Jens Hainmueller (ADH) in a series of influential papers (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015). The method is designed to estimate the causal effect of a treatment (or intervention) in settings with (at least) a single treatment unit and a number of potential control units. This is done by comparing the observed treated unit to its hypothetical trajectory in the absence of the treatment. Pre- and post-treatment data are observed for the treatment and control units (also called donors) for the outcome of interest as well as for a set of time-constant covariates. Typically, both the number of potential control units J and the number of pre-treatment periods T_{pre} are small or, at best, moderate. In many applications, T_{pre} is even smaller than J, making standard estimation approaches like Ordinary Least Squares (OLS) unreliable or infeasible.

Building on the work of Doudchenko and Imbens (2016) and ADH, we propose a variant of the regularized SC estimator (henceforth: REGSC) that shrinks individual weights toward zero and the sum of weights toward one. Due to the differentiable penalty term, the estimator has a closed-form solution, and it is flexible enough to produce reliable estimates of the counterfactual while still maintaining the interpretability of the weights. The natural Bayesian representation of our estimator is another major advantage over existing SC methods. In many applications, the estimated treatment effects are temporally cumulated over the post-treatment periods. In such cases, a simple point estimate of the effect may be misleading, and we consider it crucial to assess the estimation uncertainty in terms of uncertainty intervals. By relying on the Bayesian REGSC representation, this can be achieved via Bayesian credible intervals.

Another contribution of this chapter is to accommodate dynamic features of the time series. To achieve this, we expand the donor pool by including lagged values of both the donor series and the estimated treatment series. Such expansions often lead to a large number of additional parameters, necessitating adequate regularization. In this context, we find the

elastic net to be particularly promising due to its ability to obtain sparse solutions. Another advantage of dynamic methods is that they are well suited for dealing with possibly non-stationary time series. Note that for non-stationary but not cointegrated time series there is no stable long-run relationship between the treatment and donor series and, therefore, it does not make sense to construct a synthetic control unit by some linear combination of the donor series. In such cases, SC methods need to be applied to the differenced series. On the other hand, if the treatment series is cointegrated with the donor series, taking differences ignores important information on the long-run relationship among the series. In such a scenario, dynamic SC methods offer a significant improvement.

This chapter is structured as follows. Section 2.2 outlines the conceptual framework and motivates the REGSC estimator. The subsequent subsections bridge the gap to related estimation techniques such as principal component analysis (PCA) and discuss the role of additional explanatory variables as well as the potential presence of serial dependence in the data. We also demonstrate the estimation of Bayesian credibility intervals to quantify the estimation uncertainty. In Section 2.3, we conduct an extensive Monte Carlo simulation to study the performance of our proposed estimator and other routinely employed estimators in the context of SC. We pay special attention to plausible data-generating processes by considering static as well as stationary and non-stationary cointegrated processes. Section 2.4 demonstrates the external validity and reveals important differences between the estimators by revisiting the *Proposition 99* application of ADH. Section 2.5 concludes and provides an outlook on further work.

2.2 Theoretical Framework

The conceptual framework is related to the potential outcomes framework as introduced by Neyman (1923) and elaborated by Rubin (1974) and Holland (1986). Assume that we observe a collection of $J+1$ time series $Y_{j,t}$ with $j = 0, 1, \ldots, J$ and $t = 1, \ldots, T$. The unit $j = 0$ is denoted as the *treatment unit*, exposed to an intervention (treatment) in period $T_0 < T$. The other series with $j = 1, \ldots, J$ belong to the set of *donor units*. Accordingly, the temporal ordering is given by

$$\underbrace{1,\ 2,\ \ldots,\ T_0}_{T_{pre} \text{ observations}},\ \underbrace{T_0+1,\ T_0+2,\ \ldots,\ T}_{T_{post} \text{ observations}},$$

so that $T_{pre} = T_0$ periods of pre-treatment data and $T_{post} = T - T_0$ periods of post-treatment data are observed.

Following the existing literature, we rule out any anticipation effects and contamination (i.e.,

no spillovers to the donor units). As argued by Abadie et al. (2010), in the presence of anticipation effects, the date T_0 should be shifted backward until the no-anticipation assumption seems plausible. If units in the donor pool are affected by the treatment (contamination), these units should be removed from the donor pool prior to estimation. Our goal is to estimate the causal effect of the intervention without specifying its exact form. This is possible because the main goal of SC estimation is to estimate the counterfactual as precisely as possible, which is essentially a forecasting task. Formally, the treatment effect at time $t > T_0$ is given by

$$\delta_t = Y_{0,t} - Y_{0,t}^N \quad \text{for } t = T_0 + 1, \dots, T,$$

where $Y_{0,t}$ denotes the observed outcome under treatment and $Y_{0,t}^N$ denotes the (unobserved) potential outcome without treatment, henceforth referred to as the counterfactual outcome. We are interested in estimating the treatment effect efficiently. It is well known that the MSE-optimal forecast $\widehat{Y}_{0,t}^N$ of the counterfactual is given by the conditional expectation

$$\widehat{Y}_{0,t}^N = \mathbb{E}(Y_{0,t}^N \mid \mathcal{I}_t),$$

where $\mathcal{I}_t$ denotes the relevant information set in period t. In what follows, the information set is assumed to be given by

$$\mathcal{I}_t = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \dots, \boldsymbol{z}_t\},$$

where

$$\boldsymbol{z}_t = (Y_{1,t}, Y_{2,t}, \dots, Y_{J,t})'$$

collects the donor-unit outcomes at time t. The information set may be extended by including additional predictors represented by a $k \times 1$ vector $\boldsymbol{x}_k$, which are unaffected by the intervention but help predict the counterfactual outcome. We briefly discuss this extension in Section 2.2.4. Since such an extension complicates the analysis without providing substantial additional insight, our primary focus is on settings without additional covariates. We further assume that the $(J+1)$-dimensional time series vector

$$\boldsymbol{y}_t = \left(Y_{0,t}^N,\ \boldsymbol{z}_t'\right)'$$

is weakly stationary for $t = 1, \dots, T$, where $Y_{0,t}^N = Y_{0,t}$ for $t \leqslant T_0$. In Section 2.2.6, we discuss extensions to settings in which the autoregressive representation of $\boldsymbol{y}_t$ has roots on the unit circle (unit roots).

2.2.1 The Static Case

Let us first consider the simplest scenario with one treatment unit $j = 0$ and two donor units $j = 1, 2$. For $t = 1, \ldots, T_0$ we observe $\boldsymbol{y}_t = (Y_{0,t}, Y_{1,t}, Y_{2,t})'$, and for $t = T_0 + 1, \ldots, T$ the corresponding vector containing the counterfactual outcome, $\widetilde{\boldsymbol{y}}_t = (Y^N_{0,t}, Y_{1,t}, Y_{2,t})'$. We assume that

$$\widetilde{\boldsymbol{y}}_t \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

with $\boldsymbol{\mu} = (\mu_0, \mu_1, \mu_2)'$ and the positive definite block covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_0^2 & \boldsymbol{\sigma}_{12}' \\ \boldsymbol{\sigma}_{12} & \boldsymbol{\Sigma}_2 \end{pmatrix},$$

where $\sigma_0^2 = \mathrm{Var}(Y^N_{0,t})$, $\boldsymbol{\Sigma}_2$ is the (2×2) covariance matrix of $(Y_{1,t}, Y_{2,t})'$, and $\boldsymbol{\sigma}_{12}$ denotes the corresponding covariance vector. We are interested in deriving the mean-square optimal forecast of the counterfactual $Y^N_{0,t}$, which equals the conditional expectation

$$\mathbb{E}(Y^N_{0,t} \mid Y_{1,t}, Y_{2,t}) = \mu_0 + w_1(Y_{1,t} - \mu_1) + w_2(Y_{2,t} - \mu_2) = \mu^* + w_1 Y_{1,t} + w_2 Y_{2,t},$$

where $\mu^* = \mu_0 - w_1\mu_1 - w_2\mu_2$ and

$$\boldsymbol{w} = \begin{pmatrix} w_1 \\ \\ w_2 \end{pmatrix} = \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\sigma}_{12}.$$

The original SC estimator imposes the restrictions $w_1 \geqslant 0$, $w_2 \geqslant 0$, and $w_1 + w_2 = 1$, yet our result shows that in this simple Gaussian setting there is no inherent *statistical* justification for these constraints.[1] Moreover, the SC should include a constant term in order to adjust for different means of the donor series, as otherwise the estimated counterfactual may lie outside the convex hull of the donor means. See Doudchenko and Imbens (2016) for a detailed discussion. To illustrate these findings, assume that

$$\boldsymbol{y}_t \overset{iid}{\sim} \mathcal{N}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.1 & 0.4 \\ 0.1 & 1 & 0.5 \\ 0.4 & 0.5 & 1 \end{bmatrix} \right).$$

[1] On the other hand it is argued that restricting weights to be positive and sum to one yields an interpretable, convex combination of real units, avoids extrapolation outside the donor pool, and yields a stable, regularized estimator that minimizes overfitting. In this section we assume that all parameters of the statistical model were known. The case with unknown parameters is considered in Section 2.2.2.

For this example, the unrestricted optimal weights are $w_1 = -0.1333$, $w_2 = 0.467$, and $\mu^* = 0.667$. Note that w_1 is negative even though all pairwise correlations are positive. The result may be seen as economically counterintuitive, as interpreting $Y_{1,t}$ entering the counterfactual $\widehat{Y}_{0,t}^N$ with a negative weight is unclear. This illustrates the trade-off between statistical optimality and economic interpretability. If we impose the restrictions $w_1 \geqslant 0$, $w_2 \geqslant 0$, and $w_1 + w_2 = 1$, the restricted optimum yields the linear combination $\widetilde{Y}_{0,t}^N = 0.2\, Y_{1,t} + 0.8\, Y_{2,t}$. The key difference lies in the variance of the resulting estimates. For our example,

$$\begin{aligned} \mathrm{Var}(Y_{0,t}^N - \widehat{Y}_{0,t}^N) &= 0.8267, \\ \mathrm{Var}(Y_{0,t}^N - \widetilde{Y}_{0,t}^N) &= 1.1600. \end{aligned}$$

It is noteworthy that the variance of the restricted estimate exceeds the unconditional variance of $Y_{0,t}$; this is possible because $(w_1, w_2) = (0, 0)$ is excluded from the restricted parameter space. Under such conditions, it may be more efficient to dispense with donor units when constructing the counterfactual.

In microeconometric settings it is often assumed that units in the treatment and control groups are uncorrelated or even independent. In such cases the optimal weights are zero, and the optimal forecast of the counterfactual reduces to $\mu^* = \mu_0$, so the mean is estimated without using the donor units. Control units become relevant only under the additional assumption that $Y_{0,t}$ and $Y_{j,t}$ share the same mean for $j = 1, \ldots, J$, in which case observations from the donor pool can improve the estimation of μ_0.

2.2.2 Regularized Synthetic Control Estimator

In the previous section we considered a framework in which all distributional parameters were known. In empirical applications this is typically not the case, and the parameters must be estimated. In the SC context, samples are often small, as the variables of interest are often measured at quarterly or annual frequency. Hence, the number of pre-intervention periods T_0 is usually limited and may even be smaller than the number of donor units J. In such cases the unrestricted OLS estimator may suffer from instability, large variance, or may not even exist when $T_0 < J$.

Consider the least-squares estimator for the regression

$$Y_{0,t} = \mu^* + w_1 Y_{1,t} + w_2 Y_{2,t} + \cdots + w_J Y_{J,t} + u_t, \quad t = 1, \ldots, T_0.$$

Under the standard linear-regression assumptions and a *fixed* number of regressors J, the

OLS estimator $\widehat{\boldsymbol{w}} = (\widehat{w}_1, \ldots, \widehat{w}_J)'$ converges in probability to the true weight vector $\boldsymbol{w} = (w_1, \ldots, w_J)'$ and has a normal limit distribution as $T_0 \to \infty$. In practice, however, J is often comparable in magnitude to T_0. As shown by Bekker (1994), if $J/T_0 \to c > 0$ as $J, T_0 \to \infty$, the OLS estimator no longer converges to the true coefficients. A natural remedy is to introduce regularization.

In this context, Doudchenko and Imbens (2016) propose an elastic net regression that minimizes

$$Q(\boldsymbol{w}, \mu^*, \lambda_1, \lambda_2) = \sum_{t=1}^{T_0} \underbrace{\left(Y_{0,t} - \mu^* - \sum_{j=1}^{J} w_j Y_{j,t}\right)^2}_{\text{RSS}} + \lambda_1 \underbrace{\left(\sum_{j=1}^{J} w_j^2\right)}_{\text{Ridge}} + \lambda_2 \underbrace{\left(\sum_{j=1}^{J} |w_j|\right)}_{\text{Lasso}}.$$

The L_2-norm (Ridge penalty) of Hoerl and Kennard (1970) shrinks coefficients toward zero but does not perform variable selection, while offering a simple quadratic problem with a closed-form solution. In contrast, the L_1-norm (Lasso penalty) of Tibshirani (1996) induces both shrinkage and automatic variable selection, often setting some elements of $\boldsymbol{w}$ exactly to zero, though at the cost of requiring numerical optimization. The shrinkage parameters λ_1 and λ_2 are typically chosen by cross-validation.

We introduce a novel regularization approach, the Regularized Synthetic Control Estimator (REGSC). This estimator augments the OLS objective by adding a Ridge penalty, together with a penalty term that shrinks the sum of coefficients toward one. The latter is closely related to priors on coefficient sums commonly used in large Bayesian vector autoregressions (see, e.g., Bańbura et al., 2010). The REGSC estimator minimizes

$$Q(\boldsymbol{w}, \mu^*, \lambda_1, \lambda_2) = \sum_{t=1}^{T_0} \underbrace{\left(Y_{0,t} - \mu^* - \sum_{j=1}^{J} w_j Y_{j,t}\right)^2}_{\text{RSS}} + \lambda_1 \underbrace{\left(\sum_{j=1}^{J} w_j^2\right)}_{\text{Ridge}} + \lambda_2 \underbrace{\left(1 - \sum_{j=1}^{J} w_j\right)^2}_{\text{weight penalty}}. \quad (2.1)$$

By combining individual shrinkage toward zero (Ridge) with joint shrinkage of the weight sum toward one, this regularization mirrors key features of both the original SC estimator and the elastic net. It preserves interpretability as weights remain close to summing to one for sufficiently large λ_2, while allowing for greater flexibility and stability in small samples. To allow for an unrestricted constant term, we demean the pre-treatment series. Let

$$\tilde{\boldsymbol{y}}_0 = \left(Y_{0,1} - \bar{Y}_0, \ \ldots, \ Y_{0,T_0} - \bar{Y}_0\right)', \qquad \bar{Y}_0 = \frac{1}{T_0} \sum_{t=1}^{T_0} Y_{0,t},$$

and define the demeaned donor matrix

$$\tilde{\boldsymbol{Z}} = \left(\boldsymbol{z}_1 - \bar{\boldsymbol{z}},\ \ldots,\ \boldsymbol{z}_{T_0} - \bar{\boldsymbol{z}}\right)', \qquad \bar{\boldsymbol{z}} = \left(\bar{Y}_1, \ldots, \bar{Y}_J\right)',$$

where $\bar{Y}_j = T_0^{-1} \sum_{t=1}^{T_0} Y_{j,t}$ and $\boldsymbol{z}_t$ denotes the vector of donor observations at time t. Minimizing $Q(\boldsymbol{w}, \mu^*, \lambda_1, \lambda_2)$ yields the closed-form estimator

$$\widehat{\boldsymbol{w}}(\lambda_1, \lambda_2) = \left(\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}} + \lambda_1 \boldsymbol{I}_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J'\right)^{-1} \left(\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{y}}_0 + \lambda_2 \mathbf{1}_J\right),$$

where $\boldsymbol{I}_J$ is the $J \times J$ identity matrix and $\mathbf{1}_J$ is a J-vector of ones. In Appendix A.1, we show that for $\lambda_1 \to \infty$ and $\lambda_1/\lambda_2 \to \phi$, the weights converge to $1/(J+\phi) \approx 1/J$ for large J, which provides a more reasonable shrinkage target than zero, as in the elastic net. As with the elastic net, the shrinkage parameters λ_1 and λ_2 can be chosen via cross-validation, where our experience suggests that λ_2 is typically substantially larger than λ_1. The REGSC method is particularly well suited for the low-frequency macroeconomic context of synthetic control, as it offers a closed-form solution and tunable hyperparameters. It produces flexible weights that are typically positive and can be reliably estimated in small samples.[2]

Note that it is also possible to include common time effects to accommodate a parallel-trends assumption. In this case, the only adjustment needed in the objective function (2.1) is to replace the original outcomes $Y_{j,t}$ by their demeaned and detrended versions

$$Y_{j,t}^{\mathrm{FE}} = Y_{j,t} - \bar{Y}_j - \bar{Y}_t + \bar{Y},$$

where

$$\bar{Y}_j = \frac{1}{T_0}\sum_{t=1}^{T_0} Y_{j,t}, \qquad \bar{Y}_t = \frac{1}{J+1}\sum_{j=0}^{J} Y_{j,t}, \qquad \bar{Y} = \frac{1}{(J+1)T_0}\sum_{j=0}^{J}\sum_{t=1}^{T_0} Y_{j,t},$$

for $j = 0, \ldots, J$ and $t = 1, \ldots, T_0$. The matrix $\tilde{\boldsymbol{Z}}$ is then constructed from the transformed donor series $Y_{j,t}^{\mathrm{FE}}$ in place of $Y_{j,t}$.

Our regularized estimator also admits a natural Bayesian interpretation. As shown in Appendix A.2, the REGSC estimator corresponds to a prior of the form

$$\boldsymbol{w} \sim \mathcal{N}(\mu_0 \mathbf{1}_J,\ \boldsymbol{V}_0), \qquad \mu_0 = \frac{\lambda_2}{\lambda_1 + J\lambda_2},$$

where $\boldsymbol{V}_0$ has diagonal elements $V_{ii} = \sigma^2(\lambda_1^{-1} - \psi/J)$ for $i = 1, \ldots, J$ and off-diagonal elements

[2] We also experimented with combining the Lasso penalty and the weight-sum penalty, but found no improvement over the proposed estimator.

$V_{ij} = -\sigma^2\psi/J$ for $i \neq j$, with

$$\psi = \lambda_1^{-1} - (\lambda_1 + J\lambda_2)^{-1}.$$

If $\lambda_1/\lambda_2 \to 0$, then $\mu_0 \to 1/J$ and $\psi \to 1$, so the prior places mass on weight vectors close to the constant vector with entries $1/J$. Moreover, if either λ_1 or λ_2 diverges, the prior becomes dogmatic. In Section 2.2.7, this Bayesian perspective is used to construct Bayesian credibility intervals.

2.2.3 Factor Models

The SC approach is typically motivated by considering a factor model of the form

$$Y_{j,t} = \mu_j + \theta_t + \boldsymbol{\lambda}_j' \boldsymbol{f}_t + u_{j,t}$$

where μ_j and θ_t are unit- and time-specific constants (see, e.g., Abadie et al., 2010; Ferman, 2021). It is assumed that the $r \times 1$ vector of common factors $\boldsymbol{f}_t$ is uncorrelated with the idiosyncratic component, and that the idiosyncratic errors are mutually and temporally uncorrelated. For simplicity, let us ignore the time- and unit-specific constants (which in practice can be replaced by their sample counterparts). We are interested in the conditional expectation:

$$\begin{aligned} \widehat{Y}_{0,t}^N &= \mathbb{E}(Y_{0,t}|Y_{1,t},\ldots,Y_{J,t}) \quad \text{for } t = T_0+1,\ldots,T \\ &= \boldsymbol{\lambda}_0' \, \mathbb{E}(\boldsymbol{f}_t|Y_{1,t},\ldots,Y_{J,t}). \end{aligned}$$

This suggests estimating the common factor as a linear combination of the donor series $\boldsymbol{z}_t = (Y_{1,t},\ldots,Y_{J,t})'$. A popular estimator with this property is the principal component (PC) estimator. The approach can easily be generalized to models with $r > 1$ factors. Let $\boldsymbol{V}_r$ denote the $J \times r$ matrix of eigenvectors associated with the r largest eigenvalues of $\boldsymbol{Z}'\boldsymbol{Z}/T_0$, where $\boldsymbol{Z} = (\boldsymbol{z}_1,\ldots,\boldsymbol{z}_{T_0})$. Accordingly, the vector of r factors for period $1 \leqslant t \leqslant T$ is estimated as $\widehat{\boldsymbol{f}}_t = \boldsymbol{V}_r'\boldsymbol{z}_t$. Furthermore, the $r \times 1$ vector of factor loadings $\boldsymbol{\lambda}_0$ is obtained from the regression

$$Y_{0,t} = \mu^* + \boldsymbol{\lambda}_0'\widehat{\boldsymbol{f}}_t + u_t \quad \text{for } t = 1,\ldots,T_0$$

Let $\widehat{\mu}^*$ and $\widehat{\boldsymbol{\lambda}}_0$ denote the OLS estimators of μ^* and $\boldsymbol{\lambda}_0$, respectively. Then the estimated counterfactual results as

$$\widehat{Y}_{0,t}^N = \widehat{\mu}^* + \widehat{\boldsymbol{\lambda}}_0' \widehat{\boldsymbol{f}}_t \ .$$

This estimator is similar to the approach proposed by Xu (2017), where our factors correspond to the "interactive fixed effects." Note that $\widehat{\boldsymbol{\lambda}}_0' \widehat{\boldsymbol{f}}_t = \widetilde{\boldsymbol{w}}' \widetilde{\boldsymbol{z}}_t$ is a particular combination of the donor observations, with $\widetilde{\boldsymbol{w}} = \boldsymbol{V}_r \widehat{\boldsymbol{\lambda}}_0$. Using this approach, the original set of donors is linearly transformed into a set of linearly uncorrelated variables known as principal components. By using r principal components that capture most of the variance, a shrinkage procedure is applied that focuses on the most important components and discards the less influential ones. Furthermore, the factor model is particularly appealing when dealing with multicollinearity, which is often the case in SC applications. In this sense, the factor model can be viewed as a regularized estimator that shrinks the weights toward the optimal values implied by the factor structure. We therefore recommend its use for estimating the counterfactual in SC applications, even in the absence of a strict factor data-generating process.

An important drawback of the factor model is that the factors are selected in such a way that they fit all variables equally well. In our application, however, it is more important to obtain a good fit for the treated variable rather than for the donor series. In order to inform the factor analysis about the target variable, a supervised factor model may be useful. For our application, the principal covariate regression (PCovR) proposed by de Jong and Kiers (1992) appears most promising. For the model with a single factor, the objective function is

$$Q_\xi(\boldsymbol{\alpha}, \beta, \boldsymbol{\lambda}, \boldsymbol{w}) = \sum_{t=1}^{T_0} \left(Y_{0,t} - \alpha_0 - \beta\, \boldsymbol{w}'\boldsymbol{z}_t\right)^2 + \xi \sum_{t=1}^{T_0} \sum_{j=1}^{J} \left(Y_{j,t} - \alpha_j - \lambda_j \boldsymbol{w}'\boldsymbol{z}_t\right)^2 ,$$

where $\boldsymbol{f}_t = \boldsymbol{w}'\boldsymbol{z}_t$ represents the common factor. The parameter ξ is a tuning parameter that controls the relative importance of the treated series. If $\xi = 0$, then the factor model is equivalent to the unrestricted OLS estimator. The larger ξ is, the more important the fit of the factor becomes for the other variables. Accordingly, the parameter ξ can be seen as a regularization parameter that may be determined by cross-validation or an information criterion (Umbach, 2020).

2.2.4 Additional Covariates

The original SC approach assumes that we have available K additional time-constant covariates that characterize important features of the units. Let us represent this additional

information by the $K \times (J+1)$ matrix $\boldsymbol{X}$. The weight vector is chosen such that

$$\hat{\boldsymbol{w}}(\boldsymbol{V}) = \arg\min_{w} (\boldsymbol{x_0} - \boldsymbol{X_1}\boldsymbol{w})'\boldsymbol{V}(\boldsymbol{x_0} - \boldsymbol{X_1}\boldsymbol{w}),$$

where $\boldsymbol{x_0}$ denotes the first column of $\boldsymbol{X}$ (the characteristics of the treatment unit) and $\boldsymbol{X_1}$ is the $K \times J$ matrix of the remaining columns. Furthermore, the matrix $\boldsymbol{V}$ is a diagonal matrix of additional weights that reflect the relative importance of the corresponding characteristics for the minimization problem.

Note that our previous estimators can be seen as a special case of this estimation approach by treating the subset of pre-treatment observations $Y_{j,t}$ with $t = 1, \ldots, T_0$ as covariates. Accordingly, $K = T_0$, $\boldsymbol{x_0} = (Y_{0,1}, \ldots, Y_{0,T_0})'$, and the (t, j) element of $\boldsymbol{X_1}$ is equal to $Y_{j,t}$. Furthermore, the weight matrix $\boldsymbol{V}$ is the identity matrix.

An important drawback of the original SC approach is that, when estimating the weights, it does not exploit the (internal) information from the outcome series $Y_{j,t}$ with $j = 0, 1, \ldots, J$ and $t = 1, \ldots, T_0$. Let $\boldsymbol{Y}_{pre}$ denote the $T_0 \times (J+1)$ matrix of observations from the pre-intervention period. We can combine the internal and external information by defining the matrix $\boldsymbol{X}^+ = (\boldsymbol{X}', \boldsymbol{Y}_{pre}')'$ and consider the minimization

$$\hat{\boldsymbol{w}}(\boldsymbol{V}) = \arg\min_{w} (\boldsymbol{z}_0^+ - \boldsymbol{Z}_1^+ w)'\boldsymbol{V}(\boldsymbol{z}_0^+ - \boldsymbol{Z}_1^+ w),$$

where $\boldsymbol{z}_0^+$ denotes the first column of $\boldsymbol{Z}^+$ and $\boldsymbol{Z}_1^+$ collects the remaining columns. The upper block of the weight matrix $\boldsymbol{V}$ associated with $\boldsymbol{Y}_{pre}$ will typically be an identity matrix, whereas the weights in the lower diagonal block control the relative importance of the external information compared to the internal information from the past.

2.2.5 Dynamic Models

When modeling macroeconomic time series it is often assumed that the $(J+1) \times 1$ vector of time series $\boldsymbol{y}_t = (Y_{0,t}, \ldots, Y_{J,t})'$ can be represented by a vector autoregressive model given by

$$\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{\alpha} + \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p\boldsymbol{y}_{t-p} + \boldsymbol{u}_t \\
\boldsymbol{y}_t &= \boldsymbol{\mu} + \boldsymbol{A}(L)(\boldsymbol{y}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{u}_t
\end{aligned}$$

where $\boldsymbol{A}(L) = \boldsymbol{A}_1 + \boldsymbol{A}_2 L + \cdots + \boldsymbol{A}_p L^{p-1}$ denotes the $(J+1) \times (J+1)$ lag polynomial, $\boldsymbol{\mu} = (\mu_0, \mu_1, \ldots, \mu_J)'$ and $\mathbb{E}(u_t u_t') = \boldsymbol{\Sigma}$ is a positive definite covariance matrix. Let us derive the optimal forecast of $Y_{0,t}$ conditional on $\mathcal{I}_t = \{\boldsymbol{z}_t, \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{T_0+1}, \boldsymbol{y}_{T_0}, \ldots, \boldsymbol{y}_1\}$ for $t \geqslant T_0$. Note that we do not include $\boldsymbol{y}_{0,t}$ in the information set for $t > T_0$, since these observations

are affected by the treatment.

Denote by $\boldsymbol{Q}$ the Cholesky factor of the inverse of the covariance matrix such that $\boldsymbol{\Sigma}^{-1} = \boldsymbol{Q}'\boldsymbol{Q}$, where $\boldsymbol{Q}$ is a lower triangular matrix such that

$$\begin{aligned}\widehat{Y}_{0,t}^N &= \mathbb{E}(Y_{0,t}^N \,|\, \mathcal{I}_t) \\ &= \mu_0 + \sum_{i=1}^{J} w_i(Y_{i,t} - \mu_i) + \boldsymbol{\beta}(L)'(\boldsymbol{y}_{t-1} - \boldsymbol{\mu}) \end{aligned} \tag{2.2}$$

where $w_i = q_{i+1}/q_1$, $\boldsymbol{q} = (q_1, q_2, \ldots, q_{J+1})'$ is the first row of the matrix $\boldsymbol{Q}'$, $\boldsymbol{\beta}(L) = \boldsymbol{\beta}_1 L + \cdots + \boldsymbol{\beta}_p L^p$, and $\boldsymbol{\beta}_j = w'\boldsymbol{A}_j/w_1$. For $t \geqslant T_0$ the vector $\boldsymbol{y}_t$ results from replacing the treated series by the non-treated counterfactual $\boldsymbol{y}_t^N = (Y_{0,t}^N, Y_{1,t}, \ldots, Y_{J,t})$, where $\widehat{Y}_{0,t}^N = Y_{0,t}$ for $t \leqslant T_0$. In practice, the unknown counterfactual $Y_{0,t}^N$ needs to be replaced by the estimate $\widehat{Y}_{0,t}^N$. Accordingly the sequence $\widehat{Y}_{0,t}^N$ is obtained from a simple recursion. An important problem with the optimal solution (2.2) is that it involves $(p+1)(J+1)$ parameters which may be difficult to estimate reliably in practice. We therefore adapt some regularization for estimating the parameters. We consider three alternative regularization schemes. First, all coefficients of the dynamic model may be regularized by applying the aforementioned REGSC method as used in the static model and proposed in Section 2.2.2. Second, the REGSC weight regularization is only applied to the vector $(w_1, \ldots, w_J)$. The additional coefficients attached to the lags of the series are only subject to the Ridge penalty. Third, all coefficients of the dynamic model may be regularized by using the elastic net penalty.

2.2.6 Non-Stationary Time Series

In most empirical applications of the SC methodology, the time series exhibit prominent time trends. Therefore, it is important to consider cases where the time series are non-stationary and some of the roots of the characteristic equation, $det[A(z)] = 0$ are equal to one (unit roots). First assume that the $J+1$ time series are integrated of order one and there exists no stationary linear combination among the time series (no cointegration). In this case the SC methods considered above need to be applied to the *differences* of the time series, as otherwise the estimation suffers from the spurious regression problem. The differences are used to forecast the differenced counterfactual $\Delta Y_{0,t}^N$ yielding the change of the treatment effects $\delta_t^{\Delta} = \Delta Y_{0,t}^I - \Delta Y_{0,t}^N$ for $t > T_0$. The treatment effect of interest is then obtained by cumulating the resulting effects of the differences yielding $\overline{\delta}_t = \delta_{T_0+1}^{\Delta} + \cdots + \delta_t^{\Delta}$.

Next, assume that the donor set and the treated variable share a common stochastic trend. This is the most favorable situation that was also considered by Harvey and Thiele (2020).

In this case there exist J independent cointegration vectors that render a stationary forecast error of the counterfactual series. We may therefore apply the same estimation procedures in Section 2.2.2 that were originally developed for stationary time series. We also note that if the series are cointegrated with the treated series, then the weights $w_1, \ldots, w_J$ can be estimated super-consistently (that is with the convergence rate $1/T$) by using the error correction representation

$$Y_{0,t} = \mu_0 + \sum_{i=1}^{J} w_i(Y_{i,t} - \mu_i) + \boldsymbol{\gamma}(L)'\Delta \boldsymbol{y}_{t-1} + u_t, \quad t = 1, \ldots, T_0. \tag{2.3}$$

2.2.7 Credibility Intervals

An estimated treatment effect naturally calls for reporting standard errors and prediction intervals. For unbiased estimators, the computation of the associated uncertainty measures is straightforward. Penalized estimators such as REGSC, however, deliberately introduce bias in exchange for variance reduction to enhance predictive performance. As a result, conventional frequentist confidence intervals are no longer directly applicable. In the following, we demonstrate how Bayesian posterior sampling can be employed to quantify the estimation uncertainty surrounding our estimated treatment effects. As mentioned in Section 2.2.2, our proposed regularization is equivalent to imposing a multivariate normal prior distribution of the following form:

$$\begin{aligned} \boldsymbol{w} &\sim \mathcal{N}(\mu_0 \mathbf{1}_J, V_0), \\ \text{where } \mu_0 &= \lambda_2/(\lambda_1 + J\lambda_2), \end{aligned}$$

and V_0 is a matrix with variances $V_{ii} = \sigma^2(\lambda_1^{-1} - \psi/J)$ for $i = 1, \ldots, J$ and covariances $V_{ij} = -\sigma^2\psi/J$ for $i \neq j$, where $\psi = \lambda_1^{-1} - (\lambda_1 + J\lambda_2)^{-1}$, see Section 2.2.2.

To illustrate the estimation of Bayesian credibility intervals for the REGSC estimator, we simulated artificial data. We generated data for $T_{pre} = 50$ periods of pre-treatment and $T_{post} = 20$ periods of post-treatment data for a single treated unit and $J = 10$ potential donors. Among these donors, only the first five provide systematic information about the counterfactual, and the error is drawn from a standard normal distribution. To obtain the credibility interval, we assume $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ along with a weakly informative inverse-gamma distribution with hyperparameters $a = b = 0.001$ for the error variance. Combining the

REGSC prior with the error distribution and the likelihood gives us the posterior distribution:

$$p(\boldsymbol{w}, \sigma^2 \mid \lambda_1, \lambda_2) \propto \underbrace{\prod_{t=1}^{T_0} \mathcal{N}_t(0, \sigma^2)}_{\text{Likelihood}} \cdot \underbrace{\mathcal{IG}(a, b)}_{\sigma^2\text{-Prior}} \cdot \underbrace{\mathcal{N}(\mu, \Sigma)}_{\boldsymbol{w}\text{-Prior}}$$

Note again the multivariate J-dimensional nature of the REGSC prior. Due to the weight penalty, the off-diagonal elements of the covariance matrix are non-zero, and the J-dimensional distribution is not equivalent to the product of J individual priors, as would be the case for univariate shrinkage procedures like Ridge. A more detailed explanation of the Bayesian computation can be found in the appendix. Note that if the error variance σ^2 is known, the REGSC estimator is directly accessible via the following analytical form:[3]

$$\begin{aligned}
\boldsymbol{w} &\sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \\
\mu &= \left(\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}} + \lambda_1 \boldsymbol{I}_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J'\right)^{-1} \left(\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{y}}_0 + \lambda_2 \mathbf{1}_J\right), \\
\Sigma &= \sigma^2 \left(\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}} + \lambda_1 \boldsymbol{I}_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J'\right)^{-1} \tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}} \left(\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}} + \lambda_1 \boldsymbol{I}_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J'\right)^{-1}.
\end{aligned}$$

Assume that we obtained 1,000 valid draws from the marginal posterior densities of the weights. To obtain the 95% credibility interval, we discard the first and the last 25 trajectories of the counterfactual. Figure 2.1 depicts the procedure.

[3] See Appendix A.3 for the derivation of the analytical form of the REGSC estimator.

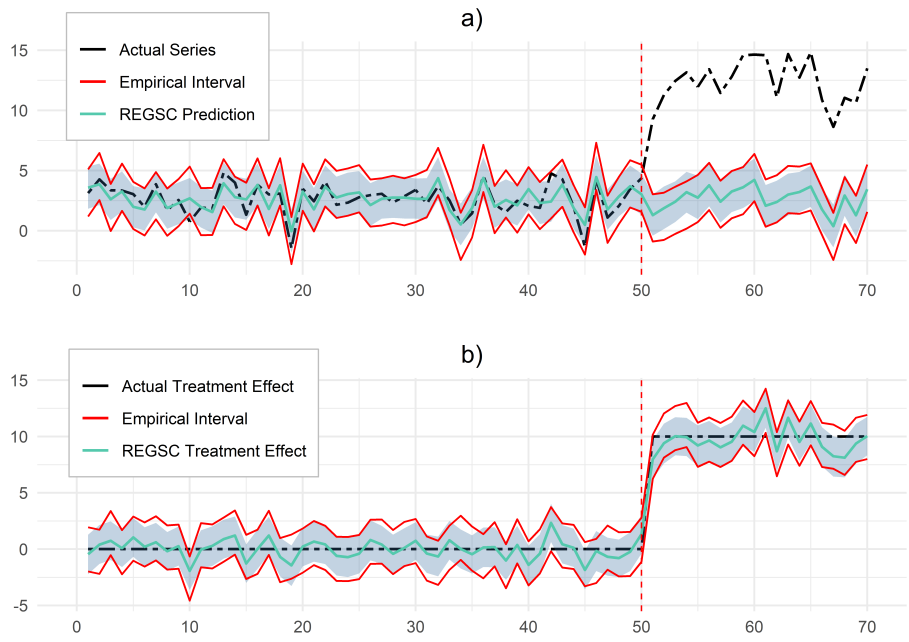

Figure 2.1: Bayesian Credibility Intervals: Panel a) plots the actual series in black, the true interval in red and the REGSC counterfactual in green. At $T = 51$, a treatment effect of size 10 is added. The Bayesian credibility interval (shaded area) provides reasonable coverage; at only two time points (10 and 45) does the actual series fall outside the shaded area. Panel b) compares the actual treatment effect (black) with the REGSC estimation (green), which is obtained by subtracting the REGSC counterfactual from the actual series. During the pre-treatment period, the actual treatment effect of zero lies within the interval in 48 out of 50 cases. In the post-treatment period, the actual treatment effect of 10 lies within the interval in 18 out of 20 cases. Furthermore, there is a high degree of similarity between the empirical confidence interval and the estimated credibility interval.

Bootstrap intervals are another way of quantifying the statistical uncertainty of the estimate. As the temporal ordering of the observations matters in applications of SC, it is advisable to perform the so-called block bootstrap, which consists of dividing the data into blocks of observations and sampling the blocks randomly with replacement (see, e.g., Hall et al., 1995).

2.3 Small Sample Performance of Alternative Estimators

In this section, we compare the performance of alternative SC estimators across different data-generating processes (DGP). We generate $T_{pre} = T_0$ periods of pre-treatment data and $T_{post} = T - T_0$ periods of post-treatment data for a single treated unit and J donor units. To align the simulation framework as closely as possible with real-world SC applications, we choose T_{pre} and T_{post} to match typical low-frequency macroeconomic settings, i.e., $T_{pre} \in$

$\{20, 50, 100\}$ and $T_{post} \in \{10\}$.[4]

We consider several types of DGPs that are stratified according to the dynamic properties of the generated data. For static data, we first examine a DGP that is specifically designed to match salient features of the SC model, followed by a static DGP based on a factor structure. For dynamic data, we begin by evaluating the models' performance under stationarity before turning to non-stationary settings with different cointegration structures.

2.3.1 Static Data

For the static data, we employ the following models:

- **SC:** The SC method of ADH with the common restrictions (no intercept, weakly positive weights that sum to one) and no additional time-invariant covariates. As outlined in Section 2.2.4, the matrix of explanatory variables $\boldsymbol{X}$ contains the pre-treatment observations $\boldsymbol{Y_{j,t}}$ of all panel units. The weights are chosen such that the pre-treatment distance between $\boldsymbol{Y_{0,t}}$ and $\boldsymbol{Y_{0,t}^N}$ is minimized. This approach to SC estimation has also been applied by Abadie et al. (2010).

- **OLS:** The usual (unrestricted) least-squares regression that regresses the pre-treatment treatment series on the donor series. The method is efficient if the number of donors (J) is small and the number of pre-treatment observations T_0 is large.

- **NET:** The elastic net as proposed by Doudchenko and Imbens (2016). The pre-treatment treatment series is regressed on the donor series with Ridge and Lasso penalties. The constant is unrestricted. Hyperparameters are chosen via cross-validation.

- **REGSC:** Our proposed regularized synthetic control estimator. The hyperparameters are selected by applying cross-validation in a two-step random grid-search procedure.

- **AUGSC:** We include the Augmented Synthetic Control estimator with Ridge penalty as an additional benchmark. This method augments the SC approach by incorporating a Ridge-penalized outcome model to improve pre-treatment fit, following the approach of Ben-Michael et al. (2021).

- **FACTOR:** Due to its implicit regularization and the factor model underlying the second static DGP, the principal component estimator from Section 2.2.3 is a natural choice.

[4] We also considered data sets with varying post-treatment sample lengths. However, the main determinant of model accuracy is the size of the pre-treatment sample.

Case 1: SC compatible DGP

We begin by evaluating the performance of the SC method and the alternative estimators under a data-generating process specifically designed to favor the characteristics of the SC model. To this end, we generate donor weights that correspond to salient features of the original SC approach by randomly drawing $J/2$ weights from a $\mathcal{U}[0,1]$ distribution and setting the remaining $J/2$ donor weights to zero. The weights are normalized to sum to one. The donor series are obtained by generating a $(T_{pre}+T_{post}) \times J$ matrix of normal random variables with covariance matrix RR', where R is a $J \times J$ matrix of independent $\mathcal{U}[0,1]$-distributed elements. The counterfactual is generated as $Y_{0,t}^N = w'Y_{j,t}^N + \epsilon_{0,t}$, with $\epsilon_{0,t}$ drawn from a standard normal distribution. We simulate 1,000 data sets, train each method on the pre-treatment period, and evaluate predictive performance in the post-treatment period. Figure 2.2 shows the simulation results, Table A.1 in Appendix A.5 reports the detailed RMSE results.

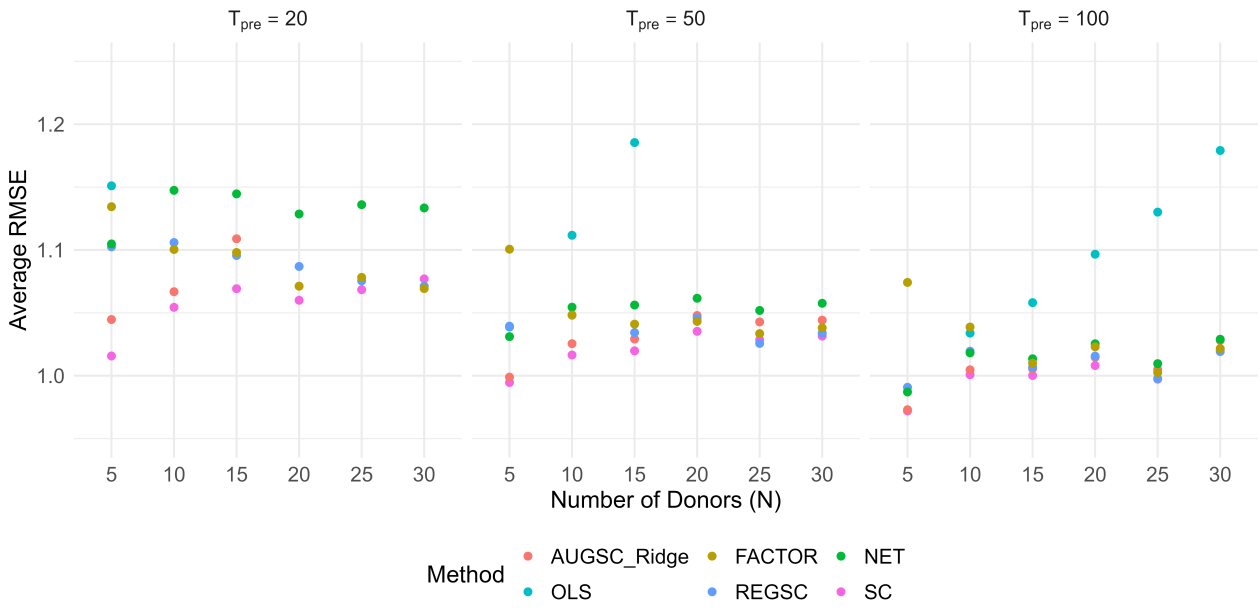

Figure 2.2: The figure shows the average RMSE of 1,000 iterations for different pre-treatment periods, donor pool sizes and methods. The y-axis range is restricted to highlight differences among the four best-performing models. Further information can be found in Table A.1.

The SC estimator performs best overall, which reflects that the DGP is constructed to follow the SC structure with sparse, non-negative weights. As a result, SC closely matches the true counterfactual and achieves the lowest RMSE in most (T_{pre}, J) combinations. REGSC follows as the second-best estimator. Its combined shrinkage on individual coefficients and the weight sum provides effective regularization while remaining flexible enough to approximate sparse donor relevance. This advantage becomes particularly pronounced as the donor pool

expands, where REGSC tends to outperform AUGSC. On the other hand, AUGSC performs well for small donor sets, where the Ridge penalty stabilizes estimation. With larger J, however, uniform shrinkage becomes too restrictive, leading to weaker performance compared to REGSC. The factor model performs reasonably well when T_{pre} is small, benefiting from dimensionality reduction. Since the DGP is not factor-driven, its accuracy does not improve much with additional donors or a longer pre-treatment period. NET ranks even below the factor model. Its combination of Lasso and Ridge penalties introduces bias in this SC-style setting and does not adapt well to the sparse/convex structure, resulting in consistently weaker performance. OLS performs worst. Without regularization, it deteriorates quickly as J increases and remains far below the regularized alternatives even when T_{pre} is large.

Case 2: Factor DGP

The original SC method is theoretically motivated by a factor model (Abadie et al., 2010). Ferman (2021) conducts simulation experiments using a simple two-factor DGP. Here we follow his simulation design but extend his approach by incorporating a time-invariant, unit-specific intercept to obtain a more realistic setting. Our representation of the counterfactual is given by

$$Y_{j,t}^N = \mu_j + \boldsymbol{\lambda}_j' \boldsymbol{f}_t + \epsilon_{j,t},$$

where $\boldsymbol{\lambda}_j$ is an $r \times 1$ vector of factor loadings, $\boldsymbol{f}_t$ is an unknown $r \times 1$ vector of common factors $\boldsymbol{f}_t = (f_{1,t}, \ldots, f_{r,t})'$, μ_j is the unit-specific intercept, and $\epsilon_{j,t}$ is an i.i.d. idiosyncratic shock.

Ferman considers a two-factor scenario, which we replicate by assigning a unit loading on the first factor to the treated unit and half of the donors, while the second factor loads on the remaining donors. Accordingly, $\boldsymbol{\lambda}_j = (1, 0)'$ for $j = 0, 1, \ldots, J/2$ and $\boldsymbol{\lambda}_j = (0, 1)'$ for $j = J/2 + 1, \ldots, J$. The random variables μ_j, $f_{1,t}$, $f_{2,t}$, and $\epsilon_{j,t}$ follow independent standard Gaussian distributions. The goal of each estimation method is to recover the underlying factor structure by assigning positive weights only to the first $J/2$ donors.

For the factor model setting, we follow the same simulation design as in the SC-style DGPs. For each combination of $T_{\text{pre}} \in \{20, 50, 100\}$ and donor pool sizes $J \in \{5, 10, 15, 20, 25, 30\}$, we generate 1,000 replications of the static factor process. As before, model performance is evaluated using the post-treatment RMSE. Analogously to the SC-type DGP, Figure 2.3 depicts the RMSE. The full simulation results can be found in Table A.2 in Appendix A.5.

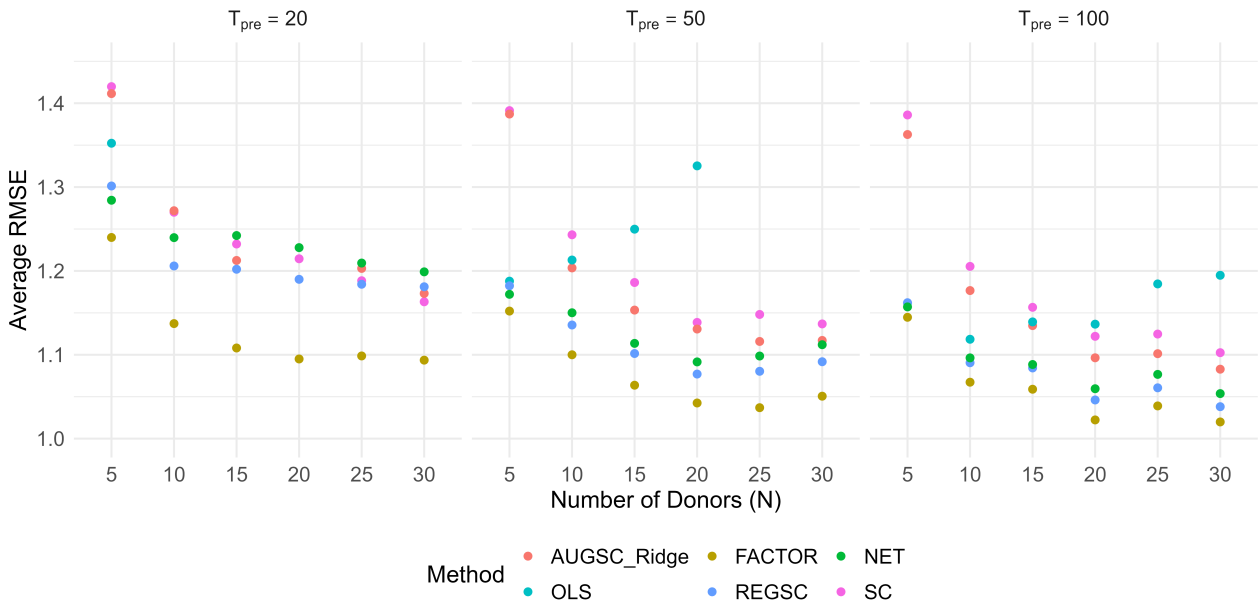

Figure 2.3: The figure shows the average RMSE of 1,000 iterations for different pre-treatment periods, donor pool sizes and methods. The y-axis range is restricted to highlight differences among the four best-performing models. Further information can be found in Table A.2.

In the factor-style DGP, all estimators improve as the pre-treatment sample and the donor pool grow, reflecting the benefits of additional cross-sectional and temporal variation. The factor model performs best in every configuration, which is expected given that the data-generating process is explicitly driven by two latent factors. Its accuracy increases steadily in T_{pre} and J, making it the clear benchmark in this setting. Among the remaining methods, REGSC and NET form the next tier of performance. Both handle the factor structure reasonably well and avoid the instability that affects unregularized approaches. REGSC performs slightly better on average (particularly for larger donor pools), since its combination of coefficient and weight-sum penalties adapts effectively to the sparse factor loading pattern implied by the DGP. NET is competitive for small donor sets but loses ground as J increases. AUGSC consistently ranks behind REGSC and NET. Its Ridge penalty stabilizes estimation in small samples, but the uniform shrinkage becomes too restrictive when the donor dimension grows, preventing it from fully capturing the factor structure. The original SC method performs relatively well in very small samples but cannot keep up once more pre-treatment periods become available. Even with larger T_{pre}, SC never outperforms REGSC or NET in this factor-driven environment. Finally, OLS is the weakest and least stable method. It is not available for $T_{pre} < J$ and, even when estimable, is prone to severe overfitting in small samples and declining accuracy as the donor pool expands.

Across the static simulation designs, several general patterns emerge. With the exception of

the unrestricted OLS estimator, all methods perform reasonably well in separating systematic pre-treatment variation from noise and show no evidence of severe overfitting. The regularized SC estimators (REGSC and NET) perform robustly across both DGPs. Their flexibility in selecting an intercept and shrinking weights helps mitigate bias and overfitting, with REGSC consistently achieving slightly lower RMSE, especially when the donor pool is large. AUGSC is reliable in small samples but falls behind REGSC and NET as J increases, reflecting the limitations of uniform Ridge shrinkage. Finally, the factor model is the strongest performer in the factor-style DGP and remains highly competitive even in the SC-style environment. Its consistently low RMSE across both simulation settings highlights the usefulness of a low-dimensional factor structure as a stable forecasting device, even when the data generate only approximate rather than exact factor relationships.

2.3.2 Dynamic Data

We now assess the performance of the previously proposed methods in a dynamic time series setting. Building on the discussion in Sections 2.2.5 and 2.2.6, which examine the implications of time series dependence and non-stationarity, we apply different estimators to a set of dynamic DGPs designed to mimic typical empirical applications. Specifically, the estimators are applied to various dynamic simulation designs, encompassing a stationary time series structure (Case 3), a non-stationary time series structure (Case 4), and a non-stationary but cointegrated time series structure (Case 5).

Case 3: Stationary DGP

In order to evaluate the performance of the alternative approaches, we proceed to study these methods in a simulation framework that mimics the real world. Given the focus of previous studies on economic development before and after treatment, it is reasonable to consider changes in GDP, such as GDP growth rates, as the basis of a realistic SC scenario. This approach anchors our Monte Carlo experiments in a realistic setting while retaining sufficient flexibility to assess estimator performance across a range of statistical scenarios. The data set assembled for this purpose comprises annual GDP growth rates of numerous countries collected over an extended period of time.[5]

The data set is restricted to countries with continuous GDP observations since 1961, ensuring a sufficiently long time series for robust simulation while preserving an adequate

[5]The GDP data are extracted from the World Bank's World Development Indicators, which is directly accessible via the WDI-Package by Arel-Bundock (2022) in R.

cross-sectional dimension. A sufficiently long time dimension is crucial for the precision of the subsequent VAR estimation, while a broad cross-section ensures an ample donor pool.[6]

Following the initial data preprocessing, a random subset of $J + 1$ countries comprising one treatment country and J potential donors is selected, and a first-order VAR model is specified and estimated. The choice to include one lag in the model is deliberate. To ensure comparability across the different cases described earlier, we employ a VAR(1) model in the stationary case, as it corresponds to a VAR(2) model with trends in the non-stationary case.[7] Consequently, we use a VAR(2) specification for Cases 4 and 5. Because the dimensionality of the VAR process grows rapidly with the lag order and the analysis targets settings with a large donor pool, we restrict the VAR specification to at most two lags to maintain computational tractability. This decision is driven by the necessity to balance model complexity and stability, and also to ensure the comparability between the cases. The estimated parameters from this estimation are used to generate the synthetic data set in our Monte Carlo experiment. The innovations are drawn from a multivariate normal distribution with a covariance matrix equal to that of the residuals from the estimated VAR model. Using the empirical residual covariance structure preserves the cross-sectional and temporal dependence observed in the data. In order to facilitate the estimation of the counterfactual, a single country is designated as the treatment unit, to which a constant treatment effect is applied. The resulting counterfactual is estimated based on the J donor series and compared to the original untreated series.

We use all methods from the static setting as benchmarks. In contrast to the previous specification with a fixed number of two factors, the number of factors in the factor model is now allowed to increase in order to accommodate the higher complexity introduced by the time series structure. Specifically, we set the number of factors equal to the square root of J. In addition, we introduce estimators that are tailored to capture specific dynamic features of the data. In particular, we apply the following dynamic approaches:

- **VAR:** The first dynamic estimator uses a VAR model. In this approach, the lagged donor series and lags of the treatment series are treated as additional donors and the corresponding weights are estimated by OLS. Consequently, this estimator is the best predictor of the synthetic control as T_{pre} tends to infinity. In finite samples, however, this estimator suffers from a large number of parameters and typically shows poor small-sample properties. Nevertheless, this estimator serves as a benchmark model for the dynamic case, as it closely mimics the DGP.

[6]The resulting data set comprises 64 (1961-2024) years of GDP per capita growth rates from a total of 29 countries.

[7]See Appendix A.4.

- **REGVAR** and **NETVAR**: As the VAR model involves a large number of parameters, it is important in empirical practice to apply some regularization scheme. We therefore extend the regularization from the static setup to the VAR model. The NETVAR method applies the elastic net penalty to all coefficients of the regressors, whereas the REGVAR approach applies the REGSC weight penalty only to the contemporaneous coefficients. We also examined cases in which the REGSC regularization was applied to all contemporaneous as well as all lagged coefficients. However, the results obtained were found to be highly similar in both cases. Consequently, the subsequent discussion will focus exclusively on the aforementioned specification.[8]

Figure 2.4 reports RMSE values for the different DGPs, pre-treatment period lengths, donor pool sizes, and methods. An overview of the detailed simulation results can be found in the Appendix A.6. The simulation results in terms of the average (Tables A.3–A.5) and median (Tables A.6–A.8) RMSE values are presented for $T_{\text{pre}} \in \{20, 50, 100\}$ and $T_{post} = 10$ periods over 1,000 iterations for each donor group and for the three aforementioned cases. In each iteration, a random subset of countries is sampled from the initially defined data set described above, and a VAR model is estimated to derive the coefficients and the covariance matrix of the error terms. These parameters are subsequently used to simulate a synthetic data set, within which one country is randomly designated as the treatment unit and the treatment effect is added. The proposed models are then applied to the generated data set, and the RMSE for the post-treatment period is computed for each model.[9]

The results indicate that dynamic methods generally outperform static methods within the dynamic DGP frameworks. When the available pre-treatment period is relatively short (here $T_{pre} = 20$), the original SC estimator performs strongly. With increasing T_{pre} and J, however, the regularized static estimators begin to substantially outperform the original SC method, and the dynamic regularized estimators in turn outperform the corresponding static versions. An expected yet still noteworthy finding is that in the stationary case the performance of all methods improves as the size of the donor pool J increases. Among the evaluated methods, the NETVAR approach achieves the lowest average RMSE, followed by the REGVAR approach. Furthermore, the findings underscore the importance of regularization in dynamic settings, especially as the number of potential donors increases. This trend is clearly reflected

[8] For all dynamic models, the lag order is set to the order used in the original VAR estimation, i.e. $p = 1$ in the stationary case and $p = 2$ in both the non-stationary cases, to align with the simulation settings.

[9] To ensure robustness of the simulation results against extreme outliers caused by numerical instabilities (e.g., matrix singularity in specific random draws), iterations yielding RMSE values in the poorest 5% of each method/donor amount combination were excluded before computing the average RMSE. Importantly, the qualitative ranking of estimators remains unchanged when using median RMSE instead of trimmed mean (see also Tables A.6–A.8 comparing median RMSE in Appendix A.6).

in the data as depicted in Tables A.6–A.8. For some simulation cases where J is relatively large compared to T_{pre}, VAR estimation frequently becomes unstable due to matrix singularity, leading to non-informative results. These cases are therefore excluded from the reported outputs.

The regularized dynamic estimators not only achieve the lowest root mean and median squared error values but also exhibit the smallest interquartile range of RMSE. This highlights both the superior performance and the increased stability of the dynamic estimators compared to their static counterparts, particularly as the number of donors increases. Furthermore, within the dynamic models, the elastic net proves to be the most effective regularization method in this simulation, underscoring its ability to generate sparse solutions.

Case 4: Non-Stationary data without cointegration structure

In many empirical settings, the underlying data exhibit non-stationary behavior. To assess the robustness of the proposed framework in such environments, we introduce two additional simulation scenarios based on non-stationary data. The first scenario (case 4) features a non-stationary process without cointegration, while the second (case 5) includes a cointegration structure.

To ensure comparability with the stationary benchmark described in Section 2.3.2, we draw GDP growth rates from the same underlying process but subsequently cumulate them to induce non-stationarity in levels. This procedure yields an $I(1)$ process with individual-specific stochastic trends. For estimation and simulation, we employ a VAR(2) specification in levels, supplemented with a linear trend to capture the induced long-run drift and maintain comparability across settings.[10] The innovations are assumed to be multivariate normal with zero mean, and for comparability, the covariance matrix is calibrated to the residual covariance matrix of the corresponding VAR model in differences. For RMSE computation, the actual and predicted counterfactual series are differenced to enable a direct comparison with the stationary results presented in Appendix A.6.

The results demonstrate that especially the static approaches face significant challenges when applied to non-stationary time series, particularly in scenarios involving larger sets of potential donors. In comparison to the (more appropriate) use of differenced time series discussed in Section 2.3.2, the RMSE is considerably higher across the board. This pronounced deterioration aligns with the implication of spurious regression. The inflated RMSEs observed

[10]Formally, if the data-generating process in first differences follows a stationary VAR(1), its accumulation implies a VAR(2) representation in levels. See Appendix A.4 for a formal derivation of the VAR(2) representation.

especially for the static specifications are fully consistent with such spurious relationships and tend to intensify as the dimensionality of the donor pool increases. The difference between static and dynamic methods becomes especially apparent when comparing the models in the setting of $T_{pre} = 100$, i.e., a setting with a comparably large period of available data. In those cases the RMSE declines for the dynamic methods with an increasing amount of donors, while it increases for static methods. This is attributable to the fact that dynamic approaches incorporate lagged differences, enabling the autoregressive specification to accommodate unit roots when necessary.

These findings underscore the importance of assessing time series' stationarity before applying static methods in standard SC settings. Particularly because differencing the time series is not always feasible or economically meaningful, particular caution is warranted when employing static methods. For cases with a large number of potential donors and sufficiently long pre-treatment periods, dynamic models (particularly the regularized REGVAR and NETVAR) are strongly recommended due to their demonstrated robustness against non-stationarity as well as their capacity to effectively handle large data sets, especially when including variable lags.

Case 5: Non-Stationary data with cointegration structure

In this subsection, we analyze the performance of the estimator when the treated and donor series are cointegrated. Economic theory and previous methodological work (see Section 2.2.6; Harvey and Thiele, 2020) indicate that in the presence of cointegration, static synthetic control methods benefit from the existence of a stable long-run relationship between treatment and donor units. This structure allows estimation on the level data without risking spurious regression, unlike in the purely integrated (non-cointegrated) case 4.

To approximate empirical scenarios as closely as possible, cointegrated data are generated by calibrating an error-correction VAR model of the form:

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 t + \Pi \mathbf{y}_{t-1} + \Gamma_1 \Delta \mathbf{y}_{t-1} + \mathbf{u}_t, \tag{2.4}$$

where the error covariance structure matches that used in the stationary and integrated cases. The $(J+1) \times (J+1)$ matrix Π is constructed to have rank J, implying a single common stochastic trend across series. This implies that the cointegrating relationship dominates long-run behavior. As in case 4, this setup ensures the comparability of the results with the other cases.

Our findings suggest that, relative to the non-cointegrated scenario, static estimators (SC,

REGSC, NET) achieve substantial improvements in predictive accuracy, with average RMSE reductions up to approximately 47% for larger donor configurations. For cases with a relatively small donor pool, the classical SC method again performs comparably well, highlighting the existence of a long-run relationship among all series. Still, dynamic approaches consistently achieve the lowest RMSE. This is due to the fact that they are able to exploit both the long-run equilibrium relationship (which is captured by the contemporaneous weights) and the short-run adjustment dynamics (which are captured by the lag polynomial). Importantly, performance in the cointegrated case consistently lies between that observed in the purely stationary setting and the integrated, non-cointegrated case. This pattern aligns with theoretical predictions: while cointegration substantially narrows the gap between static and dynamic estimators, it does not eliminate it completely.

A key finding is that regularized dynamic methods (particularly NETVAR and REGVAR) outperform the unrestricted VAR estimator even for relatively small J, emphasizing the critical importance of regularization even when cointegrating relationships are present. In all dynamic simulation cases, the net penalty seems to best balance the estimation of both long-run equilibrium weights and short-run adjustment coefficients. Especially for increasing p and J, the ability of the elastic net to generate sparse solutions is beneficial. Thus, while cointegration substantially improves the viability of static SC methods in non-stationary settings, the evidence supports dynamic regularized approaches as the preferred methodology for counterfactual estimation in macroeconomic applications, particularly across diverse donor pool configurations.

In general, our Monte Carlo experiments based on dynamic DGPs yield four essential insights. First, dynamic regularized estimators (particularly NETVAR and REGVAR) consistently outperform their static counterparts across all dynamic settings. Second, the choice of regularization is of considerable importance: the capacity of the elastic net to induce sparsity becomes especially advantageous when the potential donor pool is large, which accounts for NETVAR's modest yet systematic edge over REGVAR. Third, the relative gains of dynamic methods over static alternatives increase with the length of the pre-treatment period. Fourth, although cointegration substantially enhances the performance of static synthetic control methods compared to non-cointegrated environments, it does not eliminate the inherent performance advantage of dynamic approaches.

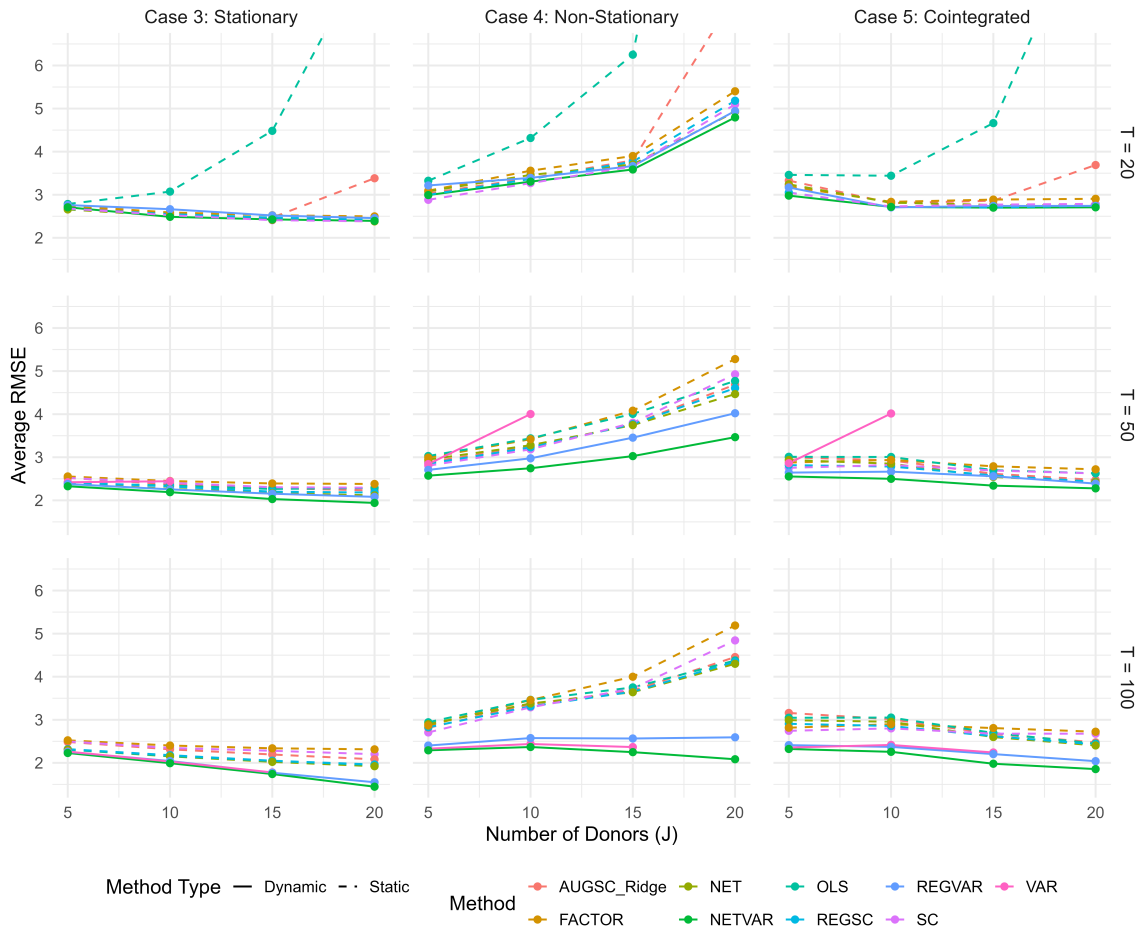

Figure 2.4: The figure shows the average RMSE of 1,000 iterations for different pre-treatment periods, donor pool sizes and methods across the three dynamic DGPs. The y-axis range is restricted to highlight differences among the four best-performing models. Further information can be found in the Tables in Appendix A.6.

2.4 Application

In this section, we examine a leading application from ADH: the effect of California's tobacco control program (Abadie et al., 2010). We compare the outcomes obtained from the different estimation methods, and pay particular attention to the prevalence of non-stationarity in the time series. For the sake of brevity, we restrict the analysis to models that performed best in the preceding simulation study. Abadie et al. (2010) estimate the effect of California's large anti-smoking legislation, *Proposition 99*. The outcome of interest is per capita cigarette consumption in California, with 38 U.S. states without similar legislation serving as donors. The authors base their analysis on $T_{pre} = 18$ (1970–1987) pre-treatment periods and $T_{post} = 13$ (1988–2000) post-treatment periods. They conclude that *Proposition 99* had a substantial and increasing negative impact on per capita cigarette sales, amounting to roughly 26 packs

per capita by the year 2000.

Figure 2.5 presents the re-estimation results for *Proposition 99* in levels and, given the strong downward trend in cigarette consumption, also in first differences. The grey area indicates the 95% credibility interval from the Bayesian representation of the REGSC model. In levels, all methods track the pre-treatment trajectory reasonably well, with predictions lying comfortably within the credibility band. A different picture emerges in first differences. The series exhibits substantial short-run volatility, with pronounced swings in both directions. While the SC estimator fails to capture this erratic behavior, NET, NETVAR, REGSC, REGVAR, and FACTOR methods handle the dynamics of the differenced data more effectively.

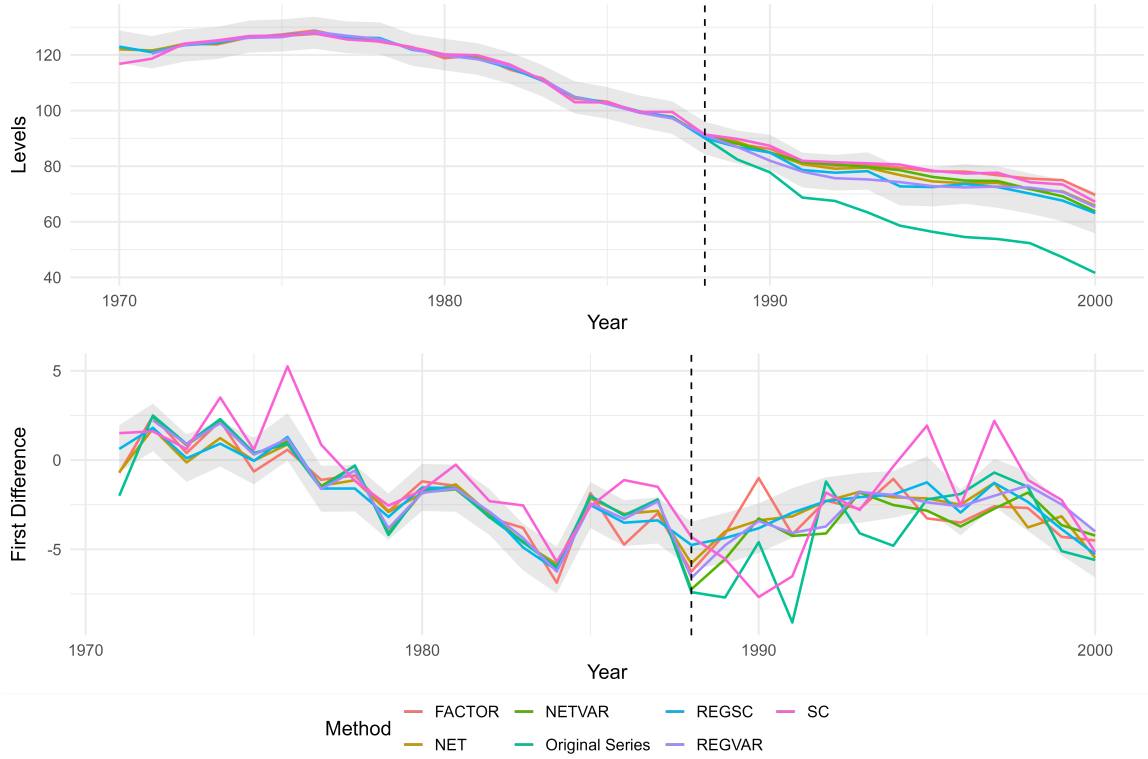

Figure 2.5: Cigarette Sales per capita for (synthetic) California. The dashed vertical line indicates the treatment date. Upper panel: original series in levels. Lower panel: transformed series of first differences.

Turning to the post-treatment period in levels, all models yield broadly similar treatment effects, with yearly averages ranging from about −14.4 (REGVAR) to −19.7 (NETVAR) packs per capita. Cumulated over the post-treatment horizon, this corresponds to reductions between roughly −173 and −237 packs per capita attributable to *Proposition 99*. The 95% REGSC credibility intervals indicate that the total effect could plausibly lie between −116 and −282. In first differences, we observe a clear deceleration in the growth rate of per-capita cigarette sales following the policy. NET, REGSC, and FACTOR produce larger but

less aligned counterfactuals than in levels, while the SC estimator struggles and fails to deliver a credible counterfactual path in either the pre- or post-treatment period.

Table A.9 in Appendix A.7 reports the full estimation results for all models and both data representations. In addition to pre-treatment R^2, end-of-sample effects (δ_{end}), mean effects ($\bar{\delta}$), and cumulative effects ($\sum \delta$), we also present the permutation p-values from ADH, computed as the ratio of post- to pre-treatment RMSE. Scaling by pre-RMSE helps avoid overly conservative test statistics by reducing the influence of donor units that are difficult to fit in the pre-treatment period. For comparability across specifications, the effects from differenced data are cumulated and transformed back to levels.

For the level data, the four static models yield plausible permutation p-values of around 5%. By contrast, the two dynamic models produce noticeably higher p-values. This pattern does not necessarily indicate model misspecification; rather, it reflects that the ADH permutation test is not well suited for more flexible estimators that achieve very accurate pre-treatment fits. When pre-treatment residuals are extremely small, even meaningful post-treatment deviations tend to be scaled down, resulting in less informative test statistics. Across all specifications based on differenced data, we find smaller and statistically less significant treatment effects. Overall, the results consistently point toward a sizeable negative impact of *Proposition 99*, although the magnitude varies across models.

2.5 Conclusion

More than twenty years ago, ADH introduced a robust and easily implemented method for estimating causal effects in macroeconomic time series. The SC approach requires no hyperparameter tuning, accommodates additional explanatory variables, and remains applicable even when the donor pool exceeds the number of pre-treatment observations. Accordingly Athey and Imbens (2016) characterize it as "arguably the most important innovation in the policy evaluation literature in the last 15 years."

The key statistical challenge in SC frameworks is to construct a reliable counterfactual estimator when the number of donors is of similar magnitude to the number of pre-treatment observations. The original synthetic control approach restricts the weights to be non-negative and to sum to unity, whereas Doudchenko and Imbens (2016) propose an elastic net regularization. Our alternative regularization instead shrinks the sum of the weights toward unity, thereby offering a more flexible version of the ADH restrictions. Motivated by the factor interpretation of SC, we additionally propose a factor-based estimator derived from a principal component analysis of the covariance matrix. Monte Carlo simulations show that both

estimators substantially improve upon existing SC methods. Moreover, the regularized SC estimator admits a natural Bayesian representation that provides credibility intervals, which are particularly useful when inference focuses on cumulative treatment effects.

Recognizing the importance of time-series dynamics in many applications, we also examine the limitations of static SC models. To address them, we propose dynamic SC estimators that augment the donor pool with lagged donor series. This greatly expands the set of predictors and therefore requires appropriate regularization to avoid overfitting. In this dynamic setting, our Monte Carlo experiments show a clear advantage of elastic net shrinkage, owing to its ability to produce sparse and stable solutions.

Another important issue is that in many empirical applications the time series (for example GDP of different countries) are non-stationary. In this case there may not exist a stable long-run relationship between the treatment and donor series that can be used to construct a reliable SC unit. Accordingly there is a risk that static SC methods suffer from some variant of the spurious regression fallacy. In this case it is important to apply the static methods to the differenced time series. The dynamic approaches circumvent this problem as they are able to accommodate unit roots in the VAR representation of the time series.

3 Fine-Tuning Expert GDP Forecasts via Dynamic Model Averaging and Google Trends

Abstract. This chapter benchmarks Dynamic Model Averaging (DMA) against professional forecasters to forecast real GDP growth in Germany from Q1-2004 to Q4-2023 using macroeconomic predictors and their corresponding Google Trends. As the informational role of Google Trends has not yet been thoroughly studied, we start with Monte Carlo simulations to systematically test their theoretical potential and find an improvement of up to 22% in forecast accuracy compared with the standard DMA model. We consider two empirical applications. First, we examine DMA's ability to outperform expert forecasts. Second, we identify the sources of discrepancies between expert forecasts and actual GDP growth, revealing that certain macroeconomic variables are either over- or underemphasized in expert forecasts. We contribute to the understanding of DMA by offering new insights into GDP dynamics, underscoring the value of integrating public perception into economic models, and understanding why experts over- or underestimate certain macroeconomic factors' relevance for GDP growth, especially during crises.

3.1 Introduction

Forecasts of key macroeconomic quantities like GDP growth are pivotal for the design of economic policy. Especially professional forecasts play a crucial role in guiding market participants and shaping their expectations about future economic conditions. Considering the political relevance of this role and the potential benefits of precise forecasting, professional forecasters strive for maximum accuracy by incorporating all the information at their disposal (Ottaviani and Sørensen, 2006). However, it is well-documented that time series models can outperform experts in forecasting (Smets et al., 2014). In this respect, the inclusion of public perception can also improve data-driven forecasting models, which can be done by means of dynamic model averaging (DMA) (Belmonte and Koop, 2014). However, whether such perception-based information truly adds value relative to model-based forecasts remains an active question. For instance, recent evidence shows that survey-based judgments can complement and sometimes outperform purely econometric forecast combinations, particularly in environments characterized by frequent structural change (Galvão et al., 2021).

Against this background, this chapter benchmarks DMA against professional forecasters to predict Germany's GDP growth (Q1-2004 to Q4-2023). On the one hand, we show that public perception enhances forecast accuracy due to better utilization of the explanatory variables instead of providing additional information. On the other hand, we unravel why professional forecasters occasionally make substantial errors in forecasting GDP growth. We use a rich set of 37 explanatory variables categorized into seven groups. The variables are supplemented by their corresponding public perception, captured through Google Trends data based on searches for each variable name and related terms. Our results show that Google Trends-supplemented DMA substantially outperforms the mean forecasts of 20 experts, especially in periods of economic uncertainty. While the use of Google Trends within the DMA framework was originally proposed by Koop and Onorante (2019), our study does not aim to introduce a new forecasting method. Rather, it builds on this existing approach and offers several distinct contributions:

- **Methodological extension:** We introduce a parsimonious modification to DMA that allows for regressor-specific coefficient variability, replacing the conventional scalar forgetting factor with a more flexible yet computationally tractable structure. This extension enables DMA to better capture heterogeneous dynamics across macroeconomic predictors and improves its stability and performance when combining variables with different sampling frequencies and persistence patterns.

- **First simulation-based evaluation of Google Trends within DMA:** We con-

duct the first systematic simulation study assessing the forecasting benefits and risks of integrating Google Trends into DMA. In contrast to previous work that used Google Trends in empirical settings, we show under controlled DGPs how Google Trends affect model averaging behavior across different signal-to-noise regimes, sample sizes, and model complexities. The simulation framework isolates the impact of each methodological component (e.g., Occam's Window, regressor-specific forgetting factors), ensuring that their effects on out-of-sample performance are not confounded by one another.

- **Empirical benchmarking against expert forecasts:** We apply the Google Trends-augmented DMA to forecast German GDP and benchmark its performance against a consensus of 20 professional forecasters. We demonstrate significant gains, especially during periods of crisis, and provide evidence that Google Trends primarily enhance information utilization rather than adding new content.

- **Forecast error diagnostics:** We propose a novel use of DMA to explain systematic expert forecast errors. By modeling the residuals between actual GDP and expert forecasts, we identify over- and underweighted macroeconomic drivers across time yielding insights into how professional forecasts could be improved.

DMA originates from Bayesian Model Averaging (BMA), which involves using the marginal likelihood of a set of candidate models as weights to estimate the relevant model quantities. Since its integration into regression, the method has found applications in various research areas, including medical research (Bahrami et al., 2023), political science (Montgomery and Nyhan, 2010), and climate modeling (Fang and Li, 2016). The two most prominent fields of application in economics include growth economics and macroeconomic forecasting. In this context, Sala-i Martin et al. (2004) employ BMA to study the determinants of long-term growth, while Koop and Potter (2004) apply BMA to forecast dynamic factor models. Although BMA has been utilized for analyzing time-varying economic data, its foundation remains rooted in a static framework (for a detailed discussion of the use of BMA in economics, we refer the reader to Steel, 2020).[1]

Raftery et al. (2010) overcome this limitation by identifying the stochastic processes of coefficients and relevant models by means of forgetting factor approximations that avoid MCMC-based inference. Thus, DMA combines the ideas of addressing model uncertainty by averaging over a series of candidate models weighted according to their a posteriori probabilities and

[1]Recent contributions also include frequentist variants of forecast combination that estimate time-varying weights without Bayesian priors (Chen and Maung, 2023). This further highlights the range of combination strategies relevant for dynamic forecasting.

of estimating dynamic coefficients (for more detailed reviews of DMA, see Nonejad, 2021; Li and Jiang, 2022). Recently, DMA has mainly been used to forecast economic quantities. Koop and Korobilis (2012) demonstrate DMA's forecasting abilities by showing that inflation predictors are subject to substantial time variations. DMA is also used for the forecasting of US house prices (Bork and Møller, 2015), crude oil prices (Naser, 2016), or stock and gold prices (Risse and Ohl, 2017). Methodologically, various extensions have been proposed. McCormick et al. (2012) consider DMA for binary classification, Onorante and Raftery (2016) incorporate the idea of Occam's Window to DMA, and Koop and Onorante (2019) elaborate on the use of external variable probabilities in the form of Google Trends. Drachal (2020) and Catania and Nonejad (2016) propose computationally efficient implementations.

Despite the numerous applications of DMA and the proposed methodological generalizations, their necessity and practical value often remain uncertain. Since this gap highlights the need for a thorough simulation study, we conduct a set of Monte Carlo simulations to determine (i) the predictive power of DMA, (ii) its ability to deal with coefficients that change dynamically over time or switch, and (iii) the (moderating) role of Google Trends. The simulations offer insights into the fine-tuning of DMA models. We find that Occam's Window has favorable regularization properties for forecasting, and we propose it as the default method. The inclusion of fully informative Google Trends leads to an improvement of about 22%. Moreover, the simulations suggest DMA as a viable alternative to recursive or rolling OLS and spline-based time-varying coefficient models.

The scope of the empirical application is two-fold. On the one hand, we start with forecasting real GDP growth and find that Google Trends-supplemented DMA yields 36% more accurate forecasts on average compared with expert forecasts (measured by RMSE). Particularly in times of crisis, our model makes substantially better forecasts than the experts do. However, our model is generally not better than the expert forecast if Google Trends are not included. Google Trends supplementation does not necessarily introduce additional information but rather enhances the utilization of information from the regressors in the model. In this respect, our empirical application shows that Google Trends have valuable moderating properties. On the other hand, we investigate how differences between expert forecasts and actual GDP growth can be explained. Within this second application, we use this difference as the dependent variable within a DMA model and pay special attention to interpreting the dynamic effects of the macroeconomic predictors, as significances indicate which variables were not sufficiently considered in the formulation of the experts' forecast. The DMA analysis shows that the experts over- and underestimate the influence of important macroeconomic factors on GDP growth at different points in time. During crises, experts often misjudged the effects of money supply, short-time employment, and stock market dynamics, overestimating

their roles. In contrast, variables such as household consumption, inflation, and business climate were underestimated in stable periods but gained prominence post-crisis. Structural changes and shocks, such as those caused by COVID-19, further highlighted mismatches in expert forecasts, particularly regarding labor volume, bank lending standards, and unemployment rates. We also explain why some of these factors are over- or underrepresented in the experts' forecasts, and discuss the policy implications.

The remainder of this chapter is organized as follows. We next outline the DMA methodology and then present the results of our simulation study. Section 3.4 contains the empirical application and Section 3.5 concludes.

3.2 Methodology

In this section, we outline the DMA framework to conduct one-step-ahead forecasts of Germany's quarterly real GDP growth in the period from 2004 to 2023. DMA is a recursive estimation technique that averages multiple time-varying parameter (TVP) regressions into a single model. It is dynamic because the coefficients in each TVP regression and the model weights vary over time. Furthermore, we (i) consider an extension that allows external variable probabilities in the form of Google Trends to determine model weights (Koop and Onorante, 2019), and (ii) the concept of Occam's Window to enable direct estimation of models with many explanatory variables in a parsimonious way (Onorante and Raftery, 2016). We also present our extension that allows for regressor-specific variability in β_t.

3.2.1 Dynamic Model Averaging

Consider the following TVP:

$$y_t = X'_{t-1}\beta_t + \epsilon_t \tag{3.1}$$

$$\beta_t = \beta_{t-1} + \eta_t, \tag{3.2}$$

where y_t is the quarterly real GDP growth for Germany, $X_{t-1} = (1, x_{1,t-1}, \ldots, x_{K,t-1})$ is a $(K+1)$-dimensional vector of intercept and K one-period lagged macroeconomic quantities. β_t is a $(K+1)$-dimensional vector of coefficients to be estimated. Both error terms are assumed iid normal, i.e., $\epsilon_t \sim \mathcal{N}(0, Q_t)$ and $\eta_t \sim \mathcal{N}(0, P_t)$. Our implementation focuses on one-step-ahead forecasts. However, multi-step forecasts can be obtained by increasing the lag between y_t and X_{t-h} for $h > 1$, leaving the recursive structure of DMA unchanged.

To begin, assume that Q_t and P_t are known. Then, the vector of time-varying coefficients β_t can be directly estimated by means of a Kalman filter (Kalman, 1960). Let $\beta_{t|t-1}$ and $\Sigma_{t|t-1}$ denote the expected value and the covariance matrix of β_t given $t-1$ information. The Kalman filter estimates for these are given by

$$\beta_{t|t-1} = \beta_{t-1|t-1} \tag{3.3}$$

$$\Sigma_{t|t-1} = \Sigma_{t-1|t-1} + P_t. \tag{3.4}$$

Based on (3.1) and (3.3), the forecast of $y_{t|t-1}$ obtains as $\hat{y}_{t|t-1} = X'_{t-1}\beta_{t|t-1}$. After time t information becomes available, coefficient vector and covariance matrix are updated as

$$\beta_{t|t} = \beta_{t|t-1} + \Sigma_{t|t-1}X_{t-1}\underbrace{(X'_{t-1}\Sigma_{t|t-1}X_{t-1} + Q_t)^{-1}}_{\text{Prediction Variance}^{-1}}\underbrace{(y_t - \hat{y}_{t|t-1})}_{\text{Prediction Error}} \tag{3.5}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1}X_{t-1}\underbrace{(X'_{t-1}\Sigma_{t|t-1}X_{t-1} + Q_t)^{-1}}_{\text{Prediction Variance}^{-1}}X'_{t-1}\Sigma_{t|t-1} \tag{3.6}$$

Estimation strategy for unknown variances

As Q_t and P_t are unknown in applied work, we follow Raftery et al. (2010) and use the idea of forgetting factors that allow the direct estimation of β_t and Σ_t for unknown Q_t and P_t. Regarding P_t, we compute

$$P_t = \left(\frac{1}{\lambda} - 1\right)\Sigma_{t-1|t-1} \tag{3.7}$$

where $\lambda \in (0, 1]$ is a forgetting factor defining the stochastic process of the covariance matrix of the state equation error term. For $\lambda = 1$, the estimation framework of (3.3) and (3.4) collapses to a linear regression with dynamic coefficients and normal errors. Values of λ below 1 imply more variation in the coefficient vector β_t as $\frac{1}{\lambda} \geqslant 1$, and result in an *age-weighted estimation* as observations i periods old receive weights λ^i. If λ is too small, the coefficient vector β_t likely becomes unstable and the estimated model is prone to overweight the most recent observations. If it is too large, the model becomes inert, making it challenging to identify true changes of β_t. To tackle this trade-off, we determine λ by computing the in-sample error over a grid of λ-values. Considering Q_t, we follow Raftery et al. (2010) and estimate Q_t by a recursive method of moments estimator that is based on the predictive distribution of the one-step-ahead forecast. As t increases, this converges to a constant.

Consider a second forgetting factor α that determines how to weight the predictions of multiple TVP regressions within a single DMA forecast. For K explanatory variables in X_{t-1} and an intercept, there exist $J = 2^K$ possible TVP regressions that constitute the DMA-forecast by means of a weighted average. To see how DMA ensures parsimony through building a weighted average, let $P(i_{t|t-1}|j_{t-1|t-1})$ denote the probability of model i in period t after model j in $t-1$ for $i, j = 1, \ldots, J$. The J model probabilities are recursively computed as

$$\pi_{t|t-1,i} = \frac{\pi^{\alpha}_{t-1|t-1,i}}{\sum_{j=1}^{J} \pi^{\alpha}_{t-1|t-1,j}}, \tag{3.8}$$

where $\alpha \in (0, 1]$ is a forgetting factor that determines how strongly contemporaneous model probabilities are impacted by past ones, i.e., how fast TVPs are allowed to switch. When y_t is observed, the model probabilities are updated as

$$\pi_{t|t,i} = \frac{\pi_{t|t-1,i} P(i_{t|t-1})}{\sum_{j=1}^{J} \pi_{t|t-1,j} P(j_{t|t-1})} \tag{3.9}$$

where $P(j_{t|t-1})$ are the predicted marginal Kalman densities, i.e., the predictive likelihood for y_t obtained by evaluating $\widehat{y}_{t|t-1}$ at the realized value for y_t. By setting a non-informative starting value of $\pi_{0|0,j} = \frac{1}{J}$, the marginal model probabilities can be recursively computed using (3.8), and updated using (3.9).

3.2.2 Extensions

Inclusion of Google Trends (Koop and Onorante, 2019)

Previous research has indicated that macroeconomic variables like inflation are of public interest (Scott and Varian, 2013, 2014). If at a specific point in time, this interest is increasing, models including this specific variable should receive higher weights than models without it. As the estimation of DMA assigns specific weights to models with different combinations of explanatory variables, external variable information reflecting their public interest likely improves the predictive power of DMA models.

We start with computing the so-called Google probability[2] of model i at time t as

$$p_{i,t} = \prod_{k \in I^i} g_{k,t} \cdot \prod_{k \in I^{\sim i}} (1 - g_{k,t}) \tag{3.10}$$

[2]Google Trends data is often referred to as 'Google probability' even though they do not fully satisfy the requirements of a probability in a true mathematical sense. For example, the axiom of additivity for disjoint events does not apply as it remains unclear what defines unrelated Google Trends. We still follow the main literature and adopt the term Google probability.

where $g_{k,t}$ represents the Google Trend of variable k at time t, scaled to range between 0 and 1, and $\sum_{i=1}^{J} p_{i,t} = 1$. I^i indicates the variables included in model i, $I^{\sim i}$ the variables not included in model i. For example, consider a DMA model with $K = 2$ and Google Trends $g_{1,t} = 0.2$ and $g_{2,t} = 0.8$. For each of the resulting $2^2 = 4$ models, the Google probabilities are solely based on the combination of probability and complementary probability of each variable:

$$
\begin{aligned}
p_{1,t} &= (1 - g_{1,t}) \cdot (1 - g_{2,t}) = 0.8 \cdot 0.2 = 0.16 \\
p_{2,t} &= (1 - g_{1,t}) \cdot g_{2,t} = 0.8 \cdot 0.8 = 0.64 \\
p_{3,t} &= g_{1,t} \cdot (1 - g_{2,t}) = 0.2 \cdot 0.2 = 0.04 \\
p_{4,t} &= g_{1,t} \cdot g_{2,t} = 0.2 \cdot 0.8 = 0.16
\end{aligned}
$$

We see that models that include variables with high Google Trends and exclude those with low Google Trends achieve higher Google model probabilities. The Google model probabilities are connected to the predictive densities of (3.8) by convex combination:

$$
\pi_{t|t-1,i}^{Google} = \omega \frac{\pi_{t-1|t-1,i}^{\alpha}}{\sum_{j=1}^{J} \pi_{t-1|t-1,j}^{\alpha}} + (1 - \omega) p_{i,t}, \tag{3.11}
$$

where $\omega \in [0, 1]$ controls the weight of the Google model probability. $\omega = 0$ implies model switching to be fully determined by Google Trends and $\omega = 1$ is the conventional DMA. After y_t has been observed, the model probabilities are again updated using (3.9).

Occam's Window

At each $t \in T$, DMA averages a total of 2^K TVP regressions. Noting that most of $t \cdot 2^K$ TVP regressions are low probability models, we consider a model subset instead of the whole space at each point in time by means of Dynamic Occam's Window (DOW) following Onorante and Raftery (2016). DOW is an approach to model selection that extends the idea of Occam's Razor by creating a window of acceptable models that balances complexity and fit. The algorithm iterates the following four steps: In the initialization step, standard DMA is performed on the subset of $K + 1$ TVPs that contain no (constant only) or exactly one explanatory variable. In the expanding step, the model population is increased by adding and subtracting one regressor to the $K+1$ models. In the forecasting step, DMA is performed on the augmented model subset. In the reducing step, only models with at least $C \cdot \pi_{t|t-1,K}^{max}$ model probability survive, where $\pi_{t|t-1,K}^{max}$ is the highest model probability and C an a priori chosen threshold. Using a set of Monte Carlo simulations (see Section 3.3), we find Occam's Window to have appealing regularization features and recommend this as the default method.

Regressor-specific state forgetting factors

In the standard DMA, λ is obtained by iterating the full estimation procedure over a grid of λ-values and choosing λ^{opt}, the forgetting factor that minimizes the in-sample error (Drachal, 2016). Hence, variation of the regressors is controlled by the covariance matrix of β_t and the scalar λ^{opt} (see equations (3.3), (3.4) and (3.7)). We propose time-invariant but regressor-specific state forgetting factors Λ, an extension that is promising whenever there is large between-coefficients variability. This allows the model to assign different levels of temporal smoothness to each regressor, which is economically meaningful: some regressors, such as long-term interest rates or demographic trends, are likely to evolve slowly over time, while others like consumer sentiment or stock returns may exhibit rapid fluctuations. Moreover, when mixing regressors sampled at different frequencies (e.g., monthly vs. quarterly), regressor-specific forgetting helps account for the heterogeneous information content and persistence across series.

As we no longer consider a scalar lambda, but a $(K+1)$-dimensional diagonal matrix Λ with regressor-specific forgetting factors on the diagonal, the approximation of P_t (compare equation (3.7)) has to be adjusted in terms of dimensionality:

$$P_t = (\Lambda^{-1} - I)\, \Sigma_{t-1|t-1} \tag{3.12}$$

In standard DMA, the optimal forgetting factor λ^{opt} is selected via the following procedure. Let l denote the number of grid points in the λ-grid, so that $l \cdot 2^K$ TVP regressions are estimated and averaged to obtain l one-step-ahead forecast sequences. For example, if the grid is $\{0.90, 0.95, 0.99\}$ (i.e., $l = 3$), three DMA models are estimated, and hence $3 \cdot 2^K$ TVP regressions. The value λ^{opt} is the forgetting factor that yields the best-performing DMA model among the candidates. In the context of regressor-specific forgetting factors, the number of TVP regressions to be estimated increases exponentially at a rate of $l^K \cdot 2^K$, making a full grid-search computationally infeasible. For example, if $l = 3$, there are already 3^K unique combinations of regressor-specific forgetting factors. For each combination, one DMA model is constructed, requiring the estimation of $3^K 2^K$ TVP regressions.

To address this, we propose selecting Λ based on individual TVP regressions rather than full DMA estimation. We start by considering the TVP regression that only includes the constant and run this regression for each of the l combinations of the λ-grid. The best-performing $\Lambda^{(0)}_{\text{opt}}$ is stored as the first element of Λ. Next, we consider the K TVP regressions that include the constant and one additional regressor, holding the forgetting factor of the constant at $\Lambda^{(0)}_{\text{opt}}$. For each of the K variables, the corresponding element of Λ is selected in the same way, by searching over the λ-grid and choosing the value that yields the best individual

forecast performance. The procedure is repeated until all K variables have appeared in a TVP including the constant and the specific regressor. The computation is parsimonious as it requires the estimation of only $l \cdot (K+1)$ TVP regressions.

The proposed extension can be summarized as follows. Allowing for regressor-specific forgetting factors provides a simple theoretical refinement: when predictors differ in persistence or respond to structural breaks at different speeds, a single scalar forgetting factor often induces either excessive smoothing or excessive volatility in some regressors. The diagonal elements of the matrix Λ allow the model to adjust the temporal smoothness of each regressor individually, better reflecting heterogeneous predictor dynamics. At the same time, the selection procedure keeps computation feasible by avoiding the exponential explosion associated with a full grid-search. This reduces the bias–variance trade-off typically present in TVP models while preserving the overall parsimony of the DMA framework.

3.3 Simulation Experiments

Simulation evidence is essential for understanding how DMA behaves under structural instability and heterogeneous predictor relevance, themes well-established in the forecast combination literature (see Hendry and Clements, 2004; Rossi, 2021). By explicitly designing DGPs with static, switching, and time-varying coefficients, we mirror the types of instabilities that are known to challenge macroeconomic forecasting models and motivate the use of combination strategies. This allows us to isolate the forecasting implications of key DMA components, including Occam's Window, forgetting factors, and the incorporation of external variable probabilities.

Since most DMA applications deal with low-frequency macroeconomic data, we consider reasonably short time spans of lengths $T \in \{60, 120, 360\}$. Our data is generated by the following dynamic regression model with time-varying coefficients for $t = 1, \ldots, T$:

$$y_t = X'_{t-1}\beta_t + \epsilon_t, \tag{3.13}$$

where β_t is the $(K+1)$-dimensional coefficient vector and X_{t-1} the $(K+1)$-dimensional vector of explanatory variables including an intercept. The variables in X_{t-1} are generated from a multivariate normal distribution with positive covariance, i.e., $X \sim MN(0, I + 0.5(\mathbf{1} - I))$, where I is an identity matrix and $\mathbf{1}$ a matrix of ones. While in our 'benchmark' case the error is $\epsilon_t \sim N(0, 0.35)$[3], we also consider different error distributions and autocorrelated errors

[3]As all columns in X_t have the same variance, the degree to which y_t can be forecasted depends only on the absolute size of the coefficients and the variance of the error term. Averaged over all iterations, the

within some robustness checks.

We begin by comparing DMA to alternative methods before we systematically assess the role of Google Trends within a second set of simulation experiments. Here, our goal is not merely to show that informative Google Trends can improve forecasts, but to quantify the magnitude of these gains, to assess the risk of deterioration when the signal is weak, and to study how DMA allocates probability mass under different regimes. In all simulations, we treat the first 30% of data as pre-sample values and consider one-step-ahead forecasts on a recursive basis. The performance of the employed models is evaluated by means of the root mean squared forecast error (RMSE) and the mean absolute forecast error (MAE). Our focus is on predictive accuracy, not inferential analysis. While formal statistical inference (e.g., confidence intervals or hypothesis testing) could be applied, we limit our evaluation to point forecast performance.

3.3.1 DMA in Comparison to Alternative Methods

To estimate the model in (3.13), we consider the following approaches: The DMA methods include variations with and without Occam's Window. The first variation incorporates Occam's Window with both a common forgetting factor λ (O_{common}; i.e., common for all coefficients, see Section 3.2.2) and a regressor-specific forgetting factor Λ ($O_{r-specific}$; i.e., allowing for heterogeneity across coefficient dimensions, see Section 3.2.2). The second variation excludes Occam's Window, using a common (NO_{common}) and a regressor-specific forgetting factor ($NO_{r-specific}$). Two methods based on ordinary least squares (OLS) are considered. The recursive OLS (recOLS) applies OLS repeatedly to an extending window of data. The rolling OLS (rolOLS) applies OLS repeatedly to a moving window of data. Additional methods include the recursive GAM (recGAM), which is a recursive additive model, where β_t is estimated using B-splines on X_{t-1}, interacted with t. The naïve forecast (naïve) predicts $y_{t+1} = y_t$.

We consider a static and a dynamic coefficient (β_t) case. Specifically, we consider none, two, and four static and dynamic coefficients β_t, i.e., $n_{stat} \in \{0, 2, 4\}$ and $n_{dyn} \in \{0, 2, 4\}$. These coefficients might change the intercept (for static coefficients) or slope (for dynamic coefficients). Furthermore, we consider up to two noise variables (regressors with corresponding coefficient of zero), i.e., $n_{noise} \in \{0, 2\}$. We limit the maximum number of columns in X_{t-1} to six for the sake of computational feasibility and use a total of 1,000 replications per case. The design of our simulation is motivated by stylized features of macroeconomic time series. Static and time-varying coefficients are included to capture structural breaks and evolving

outlined set-up yields a prediction R^2 of 0.386 for the DMA models.

relationships, as frequently discussed in the forecasting literature (see Stock and Watson, 2007; Koop and Korobilis, 2012). Noise predictors are included to reflect uncertainty about variable relevance and the risk of overparameterization in large information sets. Figure 3.1 summarizes the considered cases:

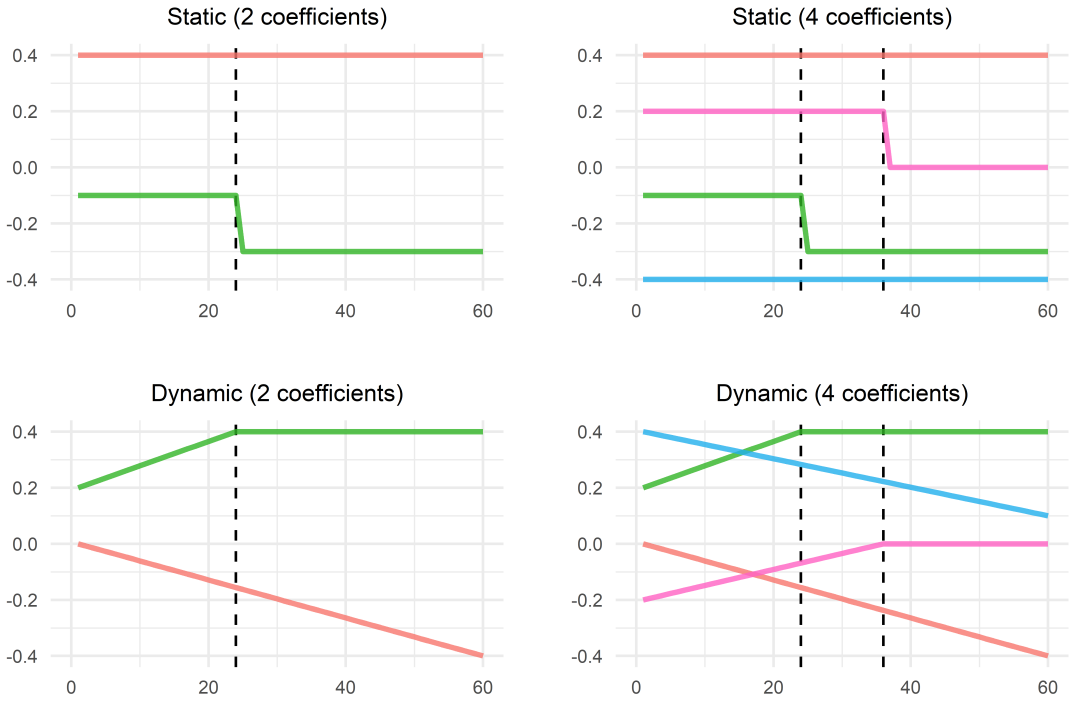

Figure 3.1: The figure shows the coefficient evolution for the static (upper row) and the dynamic (lower row) case. For both cases, sub-cases with 2 resp. 4 coefficients are considered. After 40% (24 periods) resp. 60% (36 periods), the coefficients switch.

In the 2-coefficient static case (upper-left panel), y_t is generated by a constant coefficient of 0.4 and a second static coefficient that shifts from -0.1 to -0.3 after 40% of the time span. The 4-coefficient static case (upper-right panel) introduces two additional coefficients: a constant -0.4 and a switching coefficient transitioning from 0.2 to 0 after 60% of the time span. Each case is estimated with and without noise variables, yielding four static sub-cases: $(n_{\text{stat}}, n_{\text{dyn}}, n_{\text{noise}}) \in \{(2,0,0),(2,0,2),(4,0,0),(4,0,2)\}$. The dynamic cases follow an analogous structure. In the 2-coefficient case (lower-left panel), y_t is driven by a coefficient evolving linearly from 0 to -0.4 over the full time span and another evolving linearly from 0.2 to 0.4 within the first 40%, remaining constant thereafter. The 4-coefficient scenario (lower-right panel) adds two more coefficients: one decreasing linearly from 0.4 to 0.1 over the full period and another evolving linearly from -0.2 to 0 within the first 60%, remaining uninformative for the final 40%. Noise variables are incorporated similarly to the static case,

with sub-cases $(n_{\text{stat}}, n_{\text{dyn}}, n_{\text{noise}}) \in \{(0,2,0),(0,2,2),(0,4,0),(0,4,2)\}$.

Results

Table B.1 in the appendix shows RMSE and MAE for the two static cases. recOLS is the best performing model (in terms of the RMSE) in the first sub-case (obs = 60, $n_{stat} = 2$, $n_{noise} = 0$). In all remaining 11 sub-cases, the best performing model is either O_{common} (4 times) or $O_{r-specific}$ (7 times) with a tendency of O_{common} to outperform $O_{r-specific}$ in smaller samples. Moreover, we compute the percentage by which each model's RMSE exceeds that of the best-performing model. O_{common} (+0.3%) and $O_{r-specific}$ (+0.1%) perform best, NO_{common} (+2.0%) and $NO_{r-specific}$ (+2.7%) follow. The OLS models perform substantially worse on average (recOLS: +4.7%, rolOLS: +4.4%), while recGAM (+47.1%) and the naïve model (+102.8%) are not competitive.

Table B.2 in the appendix shows RMSE and MAE for the two dynamic cases. The main results are comparable to the static case. rolOLS is the model of choice in cases with many degrees of freedom (obs = 120 and 360, $n_{dyn} = 2$, $n_{noise} = 0$). The strong link between sample size and model precision is expected, as the rolling window is defined as a fraction of the total data. Similar to the static case, O_{common} (8 times) and $O_{r-specific}$ (2 times) are the best performing models in the remaining 10 sub-cases. Also in the two sub-cases favoring OLS models, the two DMA models with Occam's Window perform, on average, just 0.6% worse than the best model. Averaged across all cases, $O_{r-specific}$ (+0.5%) and especially O_{common} (+0.1%) are highly reliable forecasting models. NO_{common} (+2.3%) and $NO_{r-specific}$ (+3.5%) fall short, as do recOLS (+5.4%) and rolOLS (+4.7%). recGAM (+35.2%) and the naïve model (+100.5%) are again worst.

As robustness checks, we also consider DGPs with (i) fat-tailed ($t(4)$), (ii) skewed (χ^2_2), and (iii) autocorrelated errors (AR(1) with $\rho = 0.6$ and $\epsilon_t = \mathcal{N}(0,1)$). We found that the DMA models are mostly robust against alternative error terms. In the context of autocorrelated errors, they are the only models to consistently improve, making them highly appropriate for applications to economic time series data.

3.3.2 DMA Models with Google Trends

The previous simulation experiments showed that the DMA specification with Occam's Window and a common forgetting factor λ (O_{common}) delivers the most reliable performance; most likely because the common forgetting factor stabilizes coefficient evolution in short samples, while Occam's Window prevents the model from allocating weight to large numbers of weak or noisy specifications. Together, these features regularize the averaging process by favoring

simpler models unless the data provide strong evidence for more complex ones. We now turn to evaluating the added value of incorporating external prior information, such as Google Trends, into the DMA framework. The key idea is that if Google Trends reflect relevant information about the underlying $x-y$ relationships, a DMA model guided by this information should outperform one that relies solely on internal model updating.

To study this systematically, we construct an artificial benchmark in which the Google Trends series are derived directly from the true coefficient paths. In this setup, each $g_{i,t}$ is chosen as a simple function of the corresponding coefficient $\beta_{i,t}$, rescaled to lie in $[0,1]$. Because Google Trends enter the DMA framework through equation (3.10), the resulting $g_{i,t}$ must generate valid model-inclusion probabilities for all 2^K candidate models. For instance, with two predictors, the pair $(g_{1,t}, g_{2,t})$ must satisfy the four probability expressions implied by (3.10). Our goal in the baseline experiment is therefore to construct $g_{i,t}$ such that these model-probability conditions are met exactly before noise is added, creating an upper-bound scenario in which Google Trends perfectly reflect the underlying relevance of the predictors.

We address the stated requirements by first assuming that models with larger coefficients in absolute value are more informative and therefore should receive higher inclusion probabilities. For each of the 2^K possible TVP regressions, we compute the sum of the absolute values of the coefficients and normalize these sums to obtain a valid probability distribution over the model space. Intuitively, a model containing two large predictors (in absolute values) will receive a higher target probability than models whose coefficients are close to zero. These target probabilities then serve as the benchmark that the Google Trends must reproduce. To obtain the corresponding K Google Trends series, we estimate them via nonlinear least squares using equation (3.10), choosing $g_{i,t}$ so that the model probabilities implied by the Google Trends closely match the target distribution. Figure 3.2 illustrates this procedure.

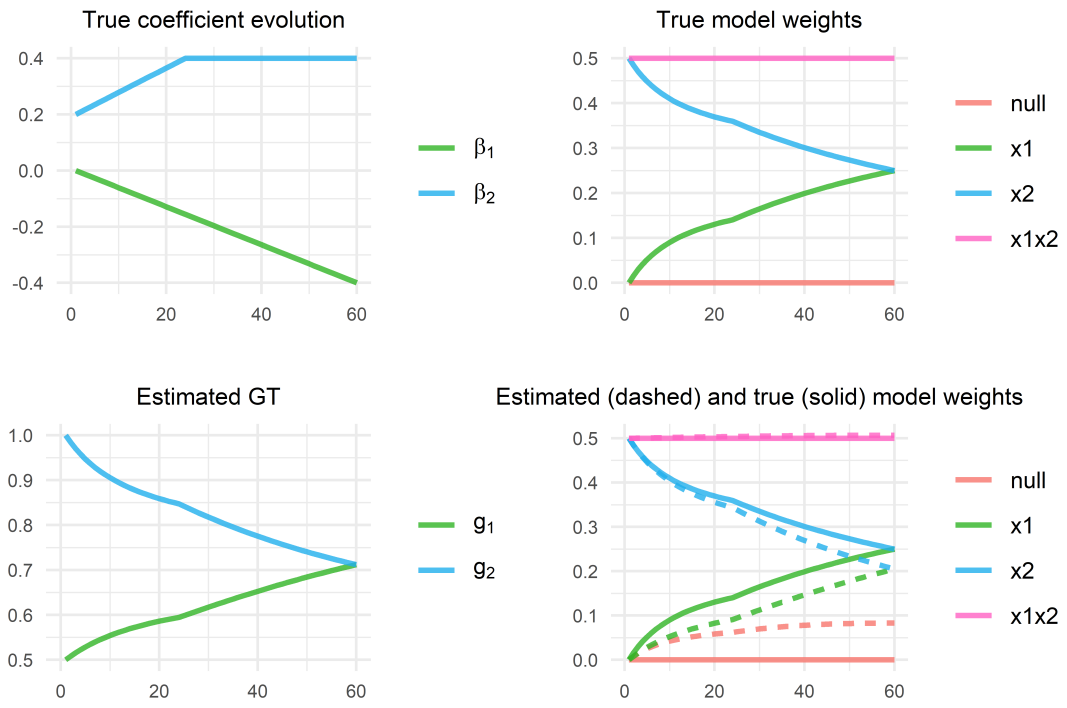

Figure 3.2: The upper left panel shows the evolution of two example coefficients (β_1 and β_2). Based on the coefficients, the upper right panel shows the optimal/true model weights. null indicates the constant-only model, x1 resp. x2 the models including only constant and x1 resp. x2. x1x2 is the full model with constant, x1 and x2. The lower left panel shows the via non-linear least squares estimated Google Trends (g_1 and g_2). The lower right panel compares optimal and (based on Google Trends) estimated model weights.

The two upper panels show how optimal model weights are computed based on prior coefficient knowledge and the relative share of the absolute coefficient sum: At the beginning of the sample, $\beta_{1,t} = 0$ and $\beta_{2,t} = 0.2$. Hence, the absolute coefficient sum of the resulting 2^2 models is 0.4 (0 for the null and the x_1-model, 0.2 for the x_2- and the x_1-x_2-model). Consequently, the null- and the x_1-model should receive weights $p_{null,t} = p_{x_1,t} = \frac{0}{0.4} = 0$. Likewise, $p_{x_2,t} = p_{x_1x_2,t} = \frac{0.2}{0.4} = 0.5$. As $\beta_{1,t}$ increases in absolute terms, the x_1-model gains and the x_2-model loses weight. The weight of the full model is constant at $\frac{1}{2^{K-1}} = 0.5$, the null model never receives a positive weight. The two lower panels show how the two Google Trends are estimated by non-linear least squares based on the 2^K possibilities of stating equation (3.10).

In our baseline scenario, we compare a standard O_{common} model to the same model estimated with perfectly informative Google Trends. This scenario represents the upper bound for the precision improvement brought about by Google Trends. To analyze the impact of noisy Google Trends, we consider cases where Google Trends provide only a weak prior signal for the x-y relationship and systematically introduce noise into the Google Trends across additional scenarios. First, we draw a $N(0, 0.01)$ noise variable, add it to the Google Trends

and truncate the sum at $[0, 1]$. The 5th and 95th percentiles of this distribution are ± 0.0164, meaning that with 90% probability, the added absolute error is below 1.64 percentage points. In scenario two, the standard deviation of the Google Trends noise term is increased to 0.1 (corresponding to a variation of ± 16.45 percentage points within the 90% range). In the last scenario, we draw the entire Google Trends from a random uniform distribution to study the risk of deterioration of fully uninformative Google Trends. We set $\omega = 0$ in equation (3.11) such that model switching is driven solely by Google Trends. Aside from the inclusion of Google Trends, all simulation settings remain unchanged from the previous section. Our results are again based on 1,000 replications per combination.[4]

Results

Table B.3 (static) and Table B.4 (dynamic) in the appendix show the results for our baseline scenario (no additional noise). Regarding the static case, we see that the Google Trends-augmented model outperforms its counterpart without Google Trends in all sub-cases by 6-22% on average. The improvement is largest when limited degrees of freedom are available and when $n_{stat} = 4$. The large improvement in sub-cases with many explanatory variables is expected since the estimation task becomes more challenging, but the Google Trends model receives more prior information. Regarding the dynamic case, we see comparable results though the improvement ranges on a smaller scale (5-17%). A possible explanation is that in the static case, Google Trends are a function of fixed coefficients, meaning the relationship with the underlying true model does not change over time.

Table B.5 (static) and Table B.6 (dynamic) show the results for the case that adds $N(0, 0.01)$ noise to the estimated Google Trends. The results are qualitatively similar to the benchmark result (improvements of 7-22% for static and 5-16% for dynamic) and we conclude that small noise in the Google Trends is negligible.

Table B.7 (static) and Table B.8 (dynamic) show the results when the variance of the Google Trends noise term increases to 0.1. This increase substantially deteriorates the signal-to-noise ratio of Google Trends. In both the static (4-16%) and the dynamic (5-16%) cases, the Google Trends model is still performing better. However, the tendency for Google Trends to be especially helpful in cases with limited degrees of freedom, as well as its greater improvement for static over dynamic coefficients, is less pronounced. This shows that even weakly informative Google Trends can enhance a DMA model. Moreover, the risk of deterioration is limited since Google Trends only guide the averaging process within DMA, while each

[4]Due to computational constraints, the results for fully uninformative Google Trends are based on only 100 repetitions.

individual model is still efficiently estimated using the Kalman filter.

To study the risk of deterioration for fully uninformative Google Trends, Table B.9 (static) and Table B.10 (dynamic) show the results for Google Trends drawn from a random uniform distribution. Compared to O_{common}, we observe deteriorations of 9-23% (static) and 7-18% (dynamic). The worst relative Google Trends performance is observed in sub-cases with 60 observations and 4 coefficients. This makes Google Trends a double-edged sword: When the signal-to-noise ratio is high, Google Trends are particularly useful for applications with limited data and many explanatory variables, some of which potentially lack explanatory power for y. However, under these same conditions, Google Trends can be most harmful when the signal-to-noise ratio is low. We conclude our simulation study with the advice to avoid setting ω in equation (3.11) too close to zero, as this would overemphasize the role of Google Trends.

3.4 Empirical Application

We present two empirical applications. In the first application, we forecast quarterly year-on-year growth of real GDP (y_t) in Germany from Q1-2004 to Q4-2023 using DMA with macroeconomic predictors and their corresponding Google Trends. The goal is to investigate to what extent Google Trends can help to outperform expert forecasts and highlight the regularization properties of Google Trends from an empirical perspective. Hence, we benchmark our model to forecasts from economic experts. We also discuss several sensitivity checks conducted to strengthen the robustness of our findings. In the second application, we analyze the accuracy of the experts' forecasts by using the difference between the true GDP growth and the mean expert forecasts ($\tilde{y}_t$) as the dependent variable in a DMA framework. For this purpose, we assume that our set of explanatory variables is a subset of the experts' explanatory variables, a weak assumption since all variables are widely recognized as being related to GDP growth. If our DMA model can account for a significant portion of the difference between the mean expert forecast and the actual growth rate, we interpret this as evidence that the experts under- or over-account for specific macroeconomic quantities.

3.4.1 Data

The two dependent variables of our two separate analyses are the quarterly year-on-year growth of real GDP (y_t) and the residual of the mean experts' forecast $\tilde{y}_t = y_t - \frac{1}{E}\sum_{e=1}^{E} Z_{t,e}$, respectively. All analyses use revised rather than real-time data, as our focus is on comparing the performance of alternative forecasting models rather than conducting a real-time fore-

casting exercise. To construct the residual $\tilde{y}_t$, we use quarterly forecasts of real GDP growth for Germany from $E = 20$ professional forecasters (experts), Z_t, who can be broadly classified into three groups: banks (11 forecasts), economic research institutes (6 forecasts), and asset managers (3 forecasts).[5] The forecasts are made at the end of each quarter and refer to GDP growth in the following quarter, i.e., they are one-quarter-ahead projections. This aligns with a realistic forecasting setup, as experts form expectations before the release of the target quarter's GDP data. Consequently, our benchmark comparison assumes that expert forecasts are made in real time, subject to the same publication delays that apply in practice. For confidentiality purposes, the actual names of the forecasters have been disguised in the subsequent analysis. Figure 3.3 shows the forecasts from the 20 experts (colored) along with the true series (black).

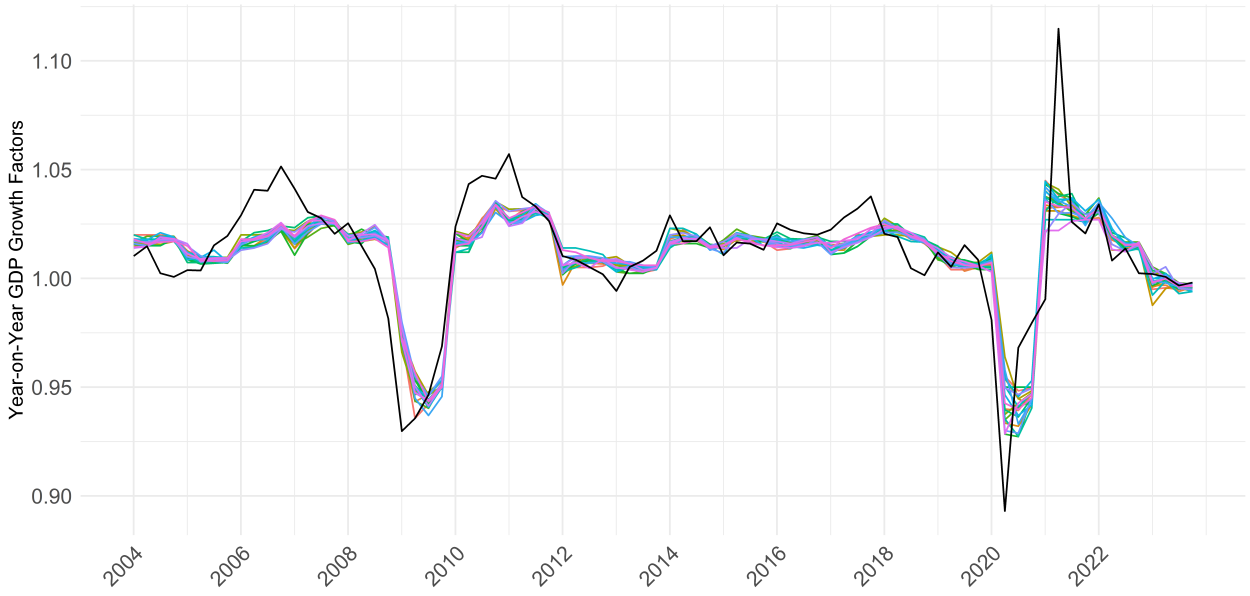

Figure 3.3: The black line shows the true GDP growth, and the colored lines show the respective expert forecasts.

We use a total of 37 macroeconomic variables in X_{t-1} and their corresponding Google Trends (p_{X_t}) as explanatory variables for y_t and $\tilde{y}_t$. The variables in X_{t-1} are grouped into seven categories: Economic Activity and Demand (EAAD, 5 variables), Labor Market and Employment (LMAE, 8 variables), Trade and External Sector (TAES, 6 variables), Financial Market Conditions (FMC, 6 variables), Price Level and Inflation (PLAI, 5 variables), Sentiments and Expectations (SAE, 3 variables), Real Estate and Housing (REAH, 4 variables)

[5]The forecasters provide on average forecasts for 83% of the 80 periods. Missing observations are filled with the mean value of the remaining forecasters at the corresponding time. Sensitivity checks confirm that our results are not driven by this approach.

(see Appendix B.3 for a complete list). We transform non-stationary variables into year-on-year growth rates.

Note that our forecasting equations do not include lagged values of y_t or $\tilde{y}_t$ as explanatory variables. While lagged dependent variables are commonly used in macroeconomic forecasting, their inclusion is inappropriate in our setting due to the use of year-on-year growth rates. Specifically, lagged values of y_t would overlap substantially with the forecast target, mechanically introducing information from the current quarter into the predictors.

We collect Google Trends data from *https://trends.google.de/trends/* by querying the name of each macroeconomic variable of interest. Each Google Trend series reports normalized search interest on a scale from 0 to 100, where 100 corresponds to peak popularity over the selected time period. To align with our framework, we rescale each series to the unit interval, allowing it to be interpreted as an external proxy for the variable's inclusion probability.

Following Koop and Onorante (2019), we enhance the signal for each macroeconomic concept by averaging the search trends of multiple related terms. While Koop and Onorante include Google's "top related searches" based on co-search frequency within user sessions, we often encountered insufficient search activity for these automated suggestions. Therefore, we manually curated additional search terms that we considered contextually relevant to each macroeconomic variable. While our selection of supplementary search terms was based on economic domain knowledge and relevant glossaries, future work may implement automated keyword extraction or taxonomy-based Google Trends construction for greater replicability. This process yields a total of 234 distinct Google Trends series. A complete list is provided in Appendix B.3.[6]

3.4.2 Forecasting GDP

In this subsection, we estimate a DMA model to forecast y_t and benchmark the resulting forecasts against the simple mean expert forecast. Given that our data set contains 37 macroeconomic predictors but roughly 70 usable quarterly observations, a direct implementation of DMA on the full variable set implies a large model space and a high risk of overfitting. To obtain a parsimonious and stable forecasting specification, we therefore extract the first principal component for each of the seven macroeconomic blocks (EAAD, LMAE, TAES, FMC, PLAI, SAE, REAH). This block-level factor structure is widely used in empirical

[6] Alternative perception measures include surveys, text-based sentiment, or uncertainty indices (see, e.g., Baker et al., 2016). We use Google Trends because it offers high-frequency, concept-specific signals that integrate naturally into DMA via variable inclusion probabilities, but the framework could accommodate other perception proxies as well.

macroeconomic forecasting, where economic information is understood to be concentrated in broad co-movement patterns within categories rather than in individual series (see, e.g., Stock and Watson, 2016). In the same vein, the corresponding Google Trends series are aggregated to the block level by averaging related search terms, ensuring consistency with the PCA-based grouping. PCA thus serves as a pragmatic dimensionality-reduction step that preserves the underlying economic grouping of variables while keeping the DMA estimation computationally feasible. Robustness checks (Section 3.4.2) confirm that the block-level PCA specification performs best among the alternatives we considered.

Table 3.1 presents the results of two DMA models, one estimated without Google Trends (columns 2-4) and one estimated with Google Trends (columns 5-7). The model with Google Trends is derived by setting $\omega = 0.5$ in equation (3.11).[7]

Table 3.1: y_t: For each model, we differentiate the outcomes for the full sample (columns 2 and 5), the financial crisis (columns 3 and 6), and the COVID-19 crisis (columns 4 and 7). The precision in terms of MAE and RMSE is presented in rows 3 and 5. In parentheses, we present the percentage performance gain of each model over the mean expert forecast. For each precision metric, we conduct Diebold-Mariano (DM) tests to assess the significance of the difference between each forecast and the mean expert forecast (rows 4 and 6). The average model size is provided in row 7. Rows 8-14 report the mean absolute size of the parameters and the number of significant quarters in parentheses. Significance is determined by constructing confidence bands assuming normally distributed errors: We compute the standard error σ_i of each DMA coefficient β_i and count how often the interval $\beta_i \pm 1.96 \cdot \sigma_i$ excludes 0. After accounting for a necessary burn-in of seven periods, the final sample contains 72 quarters (2006-Q1 to 2023-Q4). The two crises lasted 10 quarters (financial crisis, Q3-2007 to Q4-2009) and 6 quarters (COVID-19 crisis, Q1-2020 to Q2-2021), respectively.

	DMA Model			DMA-Google Trends Model ($\omega = 0.5$)		
	Full Sample	Financial Crisis	COVID Crisis	Full Sample	Financial Crisis	COVID Crisis
MAE (Gain)	0.0103 (16.94%)	0.0080 (45.95%)	0.0375 (15.54%)	0.0075 (39.52%)	0.0064 (54.73%)	0.0195 (56.08%)
DM-Test	0.1602			0.0008		
RMSE (Gain)	0.0192 (-2.13%)	0.0100 (50.74%)	0.0574 (-19.09%)	0.0120 (36.17%)	0.0089 (56.16%)	0.0299 (37.97%)
DM-Test	0.9210			0.0021		
Average Size	4.7	5.7	5.2	2.3	2.4	2.4
EAAD (n_{sig})	0.0037 (19)	0.0129 (9)	0.0030 (2)	0.0023 (21)	0.0022 (2)	0.0026 (2)
LMAE (n_{sig})	0.0062 (43)	0.0024 (3)	0.0025 (1)	0.0017 (21)	0.0002 (0)	0.0032 (3)
TAES (n_{sig})	0.0122 (55)	0.0043 (4)	0.0119 (5)	0.0085 (50)	0.0050 (4)	0.0040 (1)
FMC (n_{sig})	0.0033 (29)	0.0045 (4)	0.0001 (0)	0.0016 (11)	0.0023 (5)	0.0000 (0)
PLAI (n_{sig})	0.0028 (13)	0.0099 (7)	0.0029 (1)	0.0003 (4)	0.0010 (2)	0.0000 (0)
SAE (n_{sig})	0.0072 (26)	0.0127 (6)	0.0158 (5)	0.0053 (34)	0.0068 (3)	0.0113 (4)
REAH (n_{sig})	0.0081 (28)	0.0357 (10)	0.0030 (2)	0.0037 (20)	0.0176 (9)	0.0046 (4)

The mean expert forecast yields a MAE of 0.0124 and a RMSE of 0.0188. In comparison, the two DMA models achieve MAEs/RMSEs of 0.0103/0.0192 (without Google Trends) and

[7] We follow the recommendation of our simulation analysis and estimate the DMA models with common forgetting factor λ and conduct in-sample grid-searches to identify the optimal combination of α and λ.

0.0075/0.0120 (with Google Trends). Using the DM test, we find that for the Google Trends model, the forecast errors in terms of MAE and RMSE are significantly smaller than those of the mean expert model. For the DMA model without Google Trends, the DM test fails to reject the null hypothesis of equal forecasts between the DMA model and the mean expert forecast. A DM-test comparing the DMA model with Google Trends to the one without also indicates that forecast accuracy in terms of MAE is significantly improved by incorporating Google Trends (p-value of 0.0313). This effect is especially pronounced during the COVID-19 crisis: The main improvement is attributed to the model's precision in capturing the sharp downturn of Q2-2020 (see Figure B.1 in the appendix). Regarding the recovery after this decline, the model identifies the positive GDP development more promptly. However, we observe that the recovery effect is still somewhat underestimated by the Google Trends model.

The superior performance of the Google Trends model can be traced to how the trends enter the model probabilities in equation (3.10). In our data, the Google Trends values are generally below 0.5 (ranging from 0.1340 for SAE to 0.5244 for REAH, with an average of 0.3384). When $g_{k,t} < 0.5$, the term $(1-g_{k,t})$ dominates in equation (3.10), which mechanically increases the probability of models that exclude most predictors. As a result, the null model receives a comparatively large share of probability mass. This pattern is reflected in the much smaller average model size of the Google Trends specification (2.3) relative to the baseline DMA model (4.7). In this sense, Google Trends act as a form of regularization by encouraging shrinkage toward more parsimonious model combinations. While standard DMA can also assign high weight to the null model when supported by the data, the additional shrinkage effect induced by the Google Trends appears to strengthen this tendency. The resulting increase in null-model weighting is also visible in the smaller coefficient magnitudes shown in Figure 3.4.

In addition to introducing some coefficient variation, the coefficients in the lower panel are substantially smaller than those in the upper. As the DMA coefficients are computed as a weighted average across all 2^K models, this further indicates the Google Trends model's preference for parsimonious models. Our goal is to identify robust relationships between GDP growth and macro-categories that are not driven by methodological peculiarities. While both DMA models show similar patterns in coefficient evolution, the Google Trends model reveals more wiggly coefficients during the crises. We classify the impact of Google Trends in our application as primarily methodological rather than economic and therefore, limit the following interpretation to the model excluding Google Trends.

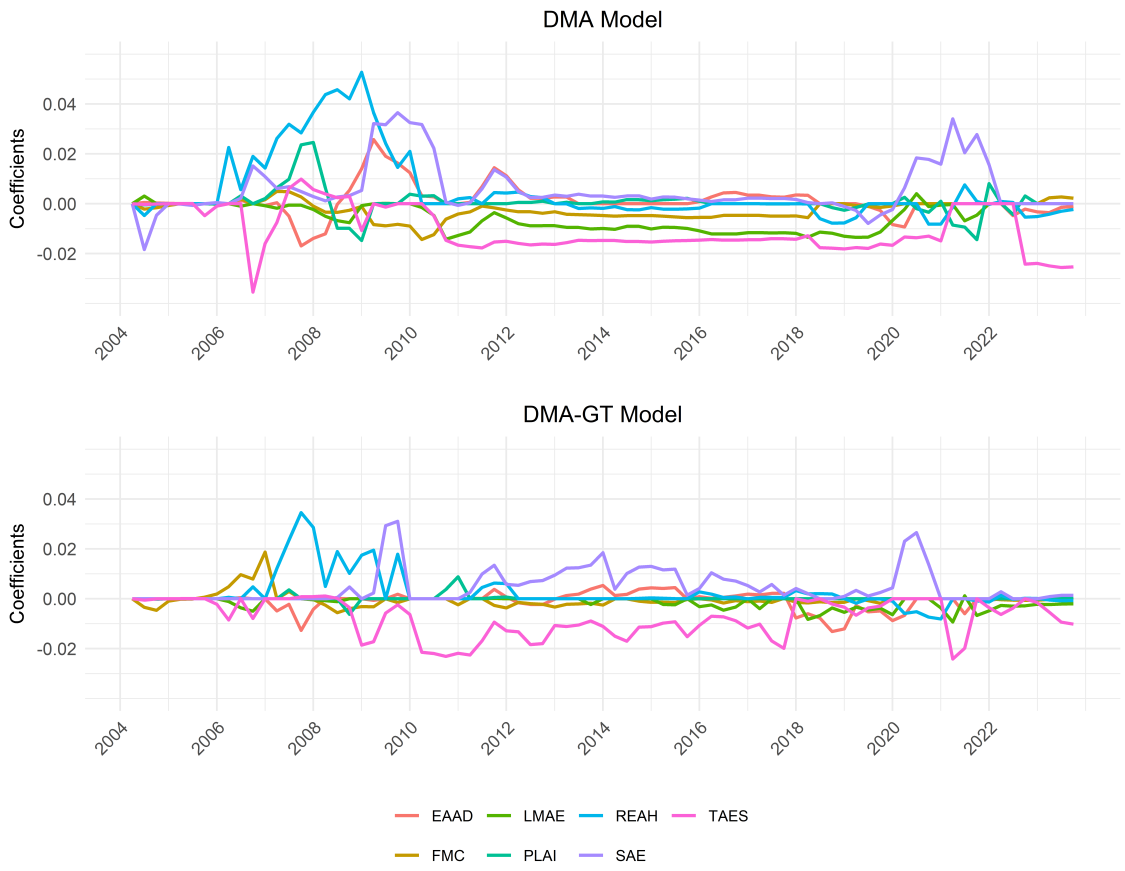

Figure 3.4: Estimated dynamic coefficients of the grouped macroeconomic variables using the DMA specification without (upper panel) and with Google Trends (lower panel).

As visible from Figure 3.4, during the financial crisis (Q3-2007 to Q4-2009), the factors relating to real estate and housing (REAH), sentiments and expectations (SAE) and economic activity and demand (EAAD) have the highest coefficients. The importance of REAH is likely a reflection of the housing market during the mortgage crisis in the U.S. The collapse of subprime lending markets and falling housing prices in the U.S. triggered global financial instability. SAE shows increased explanatory power during both crises. Consumer and firm surveys on the business climate can effectively reflect underlying economic realities, but they do not appear to serve as early indicators of economic crises, as evidenced by a coefficient that reaches its maximum only after the crisis has already begun. A similar pattern is observed during the COVID-19 crisis (Q1-2020 to Q2-2021), where we again observe the highest coefficients for SAE during the final quarter of the crisis. The second highest coefficient during the COVID-19 crisis is attached to trade and external sector (TAES). During the crisis, global supply chains were heavily interrupted. Especially for Germany and its export-oriented economy, this had serious adverse effects on the development of GDP.

Sensitivity Analysis

Several sensitivity checks were conducted, exploring both alternative models and variations in the specifications of the models presented. Regarding alternative benchmarks, we also estimated a naïve model and a DMA model with expert forecasts as explanatory variables (DMA-expert). The naïve model ($RMSE = 0.0263$) performs slightly worse, while the expert-DMA model ($RMSE = 0.0166$) performs slightly better than the mean expert model ($RMSE = 0.0188$). To ensure comparability with standard approaches used by institutions such as the Fed or the ECB, which condense their forecasts using simple averages (see, e.g., Clements et al., 2023), we chose the mean model as the benchmark.

We examined various other approaches beyond those presented. For instance, rather than incorporating Google Trends as an external probability marker, we also tried using them directly as regressors within a DMA model ($RMSE = 0.0269$). This approach was outperformed by the mean expert model, providing further evidence that, when incorporated, Google Trends should complement rather than substitute the actual variables. We also evaluated our proposed DMA model with regressor-specific forgetting factors. Consistent with the results of the simulation experiments, this version performed slightly worse than its counterpart with the common forgetting factor ($RMSE = 0.0163$). Lastly, instead of performing PCA at the category level, we also tested models that utilized factors derived directly from the ungrouped set of 37 macro variables. This approach performed significantly worse than the one employing PCA at the category level ($RMSE = 0.0202$). We conclude that the explanatory power of the factors depends not only on capturing the central dimensions of variability but also on their classification within overarching categories. Regarding alternative model specifications, we found that all presented models benefited from the extension of Occam's Window. This observation aligns with the results of the simulation experiments, which also suggested Occam's Window as the default setting. For the sake of methodological compatibility, we proceed with Occam's Window in the subsequent application.

3.4.3 Identifying Variables Overlooked in Expert GDP Forecasts

In this subsection, we examine how expert forecasts relate to macroeconomic fundamentals. Specifically, we ask whether experts systematically under- or over-weight certain macroeconomic variables when forming their expectations. To assess this, we define the dependent variable as the forecast error, $\tilde{y}_t = y_t - \frac{1}{E}\sum_{e=1}^{E} Z_{e,t}$, where y_t is actual GDP growth and $Z_{e,t}$ denotes the forecast of expert e, averaged across $E = 20$ experts. Importantly, this exercise does not attempt to replicate the real time information set of experts. While expert forecasts were made in real time-based on preliminary or incomplete data, the DMA model is

estimated using revised data and thus benefits from an informational advantage. The goal, however, is not to evaluate forecast accuracy per se, but to uncover which variables were consistently under- or over-weighted by experts in forming expectations about GDP growth.

Because our interest lies in assessing the influence of *individual* macroeconomic variables, we do not use PCA. While the main forecasting exercise employs DMA with Occam's Window to average across models, our objective here is distinct. We aim to isolate a sparse set of 'key drivers' of expert error rather than averaging over the entire probability space. Consequently, we implement a forward selection (FS) algorithm that chooses up to 10 variables out of 37 candidates. The FS procedure begins by estimating 37 separate DMA models, one per variable. The variable with the lowest RMSE in forecasting $\tilde{y}_t$ is selected. We then iteratively add variables that yield the largest RMSE improvements conditional on the variables already chosen. The process stops once no further improvement is obtained or 10 variables are selected. For technical details, see Appendix B.3.3.

To interpret the resulting coefficients, recall that the dependent variable is the forecast error $\tilde{y}_t$. We distinguish between a variable's relevance and the direction of the error: the selection of a variable by the FS algorithm serves as evidence of high relevance (analogous to a high inclusion probability), while the coefficient sign determines the direction of the weighting error. A *positive* coefficient on a macroeconomic variable X_t implies that, all else equal, experts placed *too little* weight on X_t relative to its actual influence on GDP growth. Conversely, a *negative* coefficient indicates that experts placed *too much* weight on X_t.

More formally, let $Z_{e,t} = f(X_t)$ represent expert forecasts as a function of available information. If experts fully incorporated the relevance of X_t, the sensitivity of their forecast to changes in X_t would match the true sensitivity of GDP growth, i.e., $\frac{\partial f(X_t)}{\partial X_t} = \frac{\partial y_t}{\partial X_t}$. Any divergence implies a forecast error: $\tilde{y}_t = \sum_{k=1}^{K} (\partial y_t / \partial X_{k,t} - \partial f(X_{k,t}) / \partial X_{k,t})$. Thus, the DMA coefficients capture the extent to which experts under- or overreacted to specific macroeconomic signals. A positive (negative) coefficient for $X_{k,t}$ means that its true marginal effect on GDP was underweighted (overweighted) in expert forecasts. This is not to be confused with the variable's sign of effect on GDP itself. This can be positive or negative but instead refers to the alignment (or lack thereof) between expert emphasis and actual influence.

Figures B.2 and B.3 in Appendix B.3.3 show the evolution of the forecasted GDP growth rates as well as the dynamic coefficients of the selected variables. Panel (a) in Figure B.3 displays the coefficient for money supply (M3). It is significantly negative during the financial crisis, indicating that experts *overweighted* M3 relative to its true influence. The prominence of unconventional monetary policy at the time may have contributed to this misjudgment. Between 2014 and 2018, the coefficient remains negative, again suggesting an overemphasis

on M3 in expert expectations.

The effect of the short-time volume loss (STL, panel b) was highly negative during the financial crisis but turned positive afterward, dropping to zero during COVID-19. This can be explained by short-time work that gained prominence during the financial crisis but lost its significance afterward, suggesting experts overestimated its impact during the crisis and underestimated it later. During COVID-19, the coefficient's drop to zero indicates experts may have adjusted their assessments.

Household consumption (HC, panel c) saw its coefficient rise steadily post-financial crisis, remaining significant until it dropped to zero at the onset of COVID-19. The positive trend suggests experts underestimated HC's role during stable periods. Consumer price inflation (CPI, panel d) turned significantly negative post-financial crisis, trended toward zero between 2014–2016, and became negative again just before COVID-19. During the pandemic, it shifted to significantly positive, indicating experts initially overestimated CPI's impact and underestimated it post-crisis. The German stock market index (DAX, panel e) was underestimated during the financial crisis but overestimated during COVID-19. This pattern suggests that stock markets are not consistently reliable indicators of the real economy. Dynamic coefficients are therefore needed to capture its marginal effect on GDP growth.

Labor volume (LV, panel f) was accurately accounted for until COVID-19, after which experts overestimated its impact. Post-pandemic productivity changes, such as remote work and automation, along with uneven sectoral recoveries, likely contributed to this misjudgment. The business climate (IFO, panel g) was underestimated from 2010 onward, with its coefficient peaking at the end of COVID-19. This underscores the need for greater emphasis on business climate variables, though they are not effective early indicators. Bank lending standards (BLS, panel h) were underestimated during the financial crisis and post-COVID-19 recovery. Their indirect, long-term impact on GDP, such as through investment, may have been inadequately reflected in forecasts. The unemployment rate (UR, panel i) was significantly underestimated during COVID-19, partly due to distortions from short-time work and government measures. Historical correlations between UR and GDP have likely failed to account for pandemic-specific labor market dynamics. Multiple job holders (MJ, panel j) were accurately assessed until COVID-19, after which their marginal effect was overestimated. The surge in part-time employment in low-productivity sectors and economic uncertainty during the pandemic likely weakened MJ's impact on GDP.

Since economic policy is often targeted at maintaining GDP growth and based on expert opinion, our results allow for a few shorthand conclusions for policymakers. Policymakers should rely on adaptive forecasting models that dynamically adjust to shifts in variable relationships

during crises. For monetary policy, greater emphasis should be placed on fiscal interventions targeting specific sectors instead of assuming broad effects from money supply. Labor policies must reflect changes in productivity and employment patterns, particularly during periods of structural transformation. Household consumption deserves greater attention during stable periods, as it plays a larger role in sustaining GDP growth than traditionally accounted for. Additionally, reliance on sentiment indices should be tempered by complementary data from sectoral and structural indicators to capture the complexity of economic dynamics during uncertainty. These measures will help mitigate over- or underestimations and enable more precise and effective policy responses.

3.5 Conclusion

In this chapter, we benchmark DMA against professional forecasters in predicting Germany's GDP growth (Q1-2004 to Q4-2023). Dynamic model averaging (DMA) enjoys a reputation in time series modeling (Koop and Onorante, 2019; Onorante and Raftery, 2016), especially because data about public perception can be easily supplemented (Belmonte and Koop, 2014). This perception is captured through Google Trends data, based on searches for each variable name and related terms. For this purpose, we first conduct some simulation experiments to systematically evaluate the potential of Google Trends. Specifically, we find that determining model weights based on informative Google Trends increases forecasting accuracy by up to 22% compared to standard DMA models. However, uninformative Google Trends can substantially worsen the forecasts by up to 23%.

From an empirical perspective, we show that public perception improves forecast accuracy by enhancing the utilization of explanatory variables rather than introducing new information. Our results show that most of the DMA models significantly outperform the mean expert forecast, particularly when incorporating Google Trends data as an external variable probability. Consistent with previous research (Scott and Varian, 2013; Koop and Onorante, 2019), our study provides evidence that Google Trends data should serve only to complement, rather than substitute for the actual explanatory variables by influencing the variable probability.

In a second application, we explain the differences between true GDP growth and corresponding forecasts from expert institutions using DMA. We identify substantial differences, especially during crises, which suggest over- or under-representation of certain macroeconomic variables. This analysis provides insights into which variables may require greater or lesser attention when generating GDP forecasts for Germany.

4 Panel Forecasting under Individual Heterogeneity: Estimation and Regularization Strategies

Abstract. This chapter studies forecasting in dynamic panel data models in the context of individual heterogeneity. Assuming that common parameters can be consistently estimated, the analysis focuses on recovering unit-specific intercepts for forecasting purposes. We derive optimal predictors under alternative assumptions regarding individual effects (fixed versus random) and initial conditions (stationary versus non-stationary). When initial conditions are non-stationary, the process has not yet converged and the observed data depends on the starting value, which is unknown in practice. We propose a simple decay test to identify non-stationarity in empirical practice. The proposed methods are benchmarked against plug-in approaches and a recent empirical Bayes estimator based on Tweedie's formula. Monte Carlo simulations emphasize the importance of correctly specifying initial conditions and examine the impact of additional (time-varying) covariates. It turns out that it is important to account for the correlation between individual effects and regressors using the Mundlak/Chamberlain approach (also called the correlated effects forecast). An empirical application to firm-level productivity forecasting confirms the simulation results, demonstrating that shrinkage-based predictors, particularly those designed for non-stationary initial conditions, substantially improve forecast accuracy relative to simple mean forecasts.

4.1 Introduction

The estimation of dynamic panel data models with individual effects is well understood. Nickell (1981) shows that fixed-effects estimators of dynamic models are biased in short panels, motivating alternative approaches based on orthogonality conditions or likelihood methods. Arellano and Bond (1991) propose difference generalized method of moments (GMM) estimators that eliminate unit effects by first differencing, while Blundell and Bond (1998) extend this framework by combining difference and level equations to address weak instruments. Under standard conditions, both GMM and maximum-likelihood approaches yield consistent estimates of the homogeneous parameters, such as the autoregressive coefficient and slope parameters (Moral-Benito et al., 2019). In panels with a large number of cross-sectional units (N), these common parameters can therefore be estimated precisely.

While these results establish consistent estimation of the structural parameters, forecast accuracy depends crucially on how well unit-specific components are recovered: when T is small (and thus treated as fixed in what follows), individual effects are weakly identified and their estimation error becomes a non-negligible component of the forecast error. Early work on forecasting with panel data emphasizes the potential gains from pooling information across cross-sectional units. Baltagi (2008) provides a comprehensive overview of forecasting methods for panel data models and highlights the trade-off between efficiency gains from pooling and the bias that may arise when individual specific heterogeneity is ignored (see also Pesaran et al., 2024, for a detailed comparison of both forecasting methods). In this setting, pooled forecasts may be severely biased, while fully unit-specific forecasts suffer from high estimation variance in short panels. Random-effects approaches offer a compromise by shrinking individual forecasts toward the pooled mean, thereby trading bias for variance reduction. Moreover, modeling individual effects as correlated random components, following Mundlak (1978) and Chamberlain (1982), provides a structured way to incorporate cross-sectional information and improve the estimation of unit-specific effects.

A prominent recent contribution is Liu et al. (2020), who develop an empirical Bayes approach to forecasting in dynamic panel data models. Their framework treats unit-specific effects as random and exploits the cross-sectional distribution of sufficient statistics to construct shrinkage estimators that explicitly target forecast optimality. The empirical Bayes perspective builds on the classical shrinkage literature originating with Tweedie (1957) and further developed by Efron (2011). The key intuition is that extreme estimates of individual effects tend to overstate the underlying signal because they are partly driven by idiosyncratic noise, a phenomenon commonly referred to as *the winner's curse*. The work of Liu et al. serves as the main benchmark for the analysis of this chapter.

The remainder of this chapter is organized as follows. Section 4.2 introduces the baseline dynamic panel framework without additional covariates. Section 4.2.1 discusses the role of initial conditions and proposes a simple decay test to detect non-stationarity in empirical practice. Section 4.2.2 demonstrates analytically how imprecise estimation of individual effects inflates the mean squared forecast error in short panels, thereby motivating shrinkage and regularization strategies. Section 4.2.3 constitutes the core methodological contribution of the chapter. It develops mean-squared-error (MSE)-optimal forecasting rules under alternative assumptions on individual heterogeneity and initial conditions. The section derives closed-form predictors and studies their bias–variance trade-offs. These results provide the foundation for the subsequent comparison with empirical Bayes approaches. Section 4.3 extends the framework to models with time-invariant and time-varying explanatory variables. Section 4.4 evaluates the finite-sample performance of the proposed procedures in a series of Monte Carlo experiments. The simulation results show that methods derived under non-stationary initial conditions are substantially more robust to misspecification and, in particular, that the random-effects predictor under non-stationary initialization consistently delivers the strongest forecast performance. Section 4.5 presents an empirical application to firm-level productivity forecasting based on a design-based simulation for 10,580 firms and illustrates the practical relevance of the proposed methods. Section 4.6 concludes.

4.2 Dynamic Models without Exogenous Variables

Consider the dynamic panel model

$$y_{it} = \mu_i + \rho\big(y_{i,t-1} - \mu_i\big) + \varepsilon_{it}, \qquad i = 1, \dots, N, \;\; t = 1, \dots, T, \tag{4.1}$$

where $|\rho| < 1$, $\mu_i \sim (0, \sigma_\mu^2)$ and $\varepsilon_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$. For ease of exposition, we focus on the case of a single lag and first exclude exogenous variables. To simplify notation, the unit effects are reparameterized as

$$\lambda_i := (1-\rho)\mu_i, \qquad \sigma_\lambda^2 = (1-\rho)^2\sigma_\mu^2,$$

so that μ_i represents the long-run mean and λ_i the intercept of the process. Throughout the analysis, a large-N-small-T environment is assumed. In such an environment, we can ignore the estimation error in $\hat{\rho}$ and treat ρ as known or at least estimated consistently. Note that for $\rho = 0$, the model in equation (4.1) nests the simplest possible panel model that includes only the individual effect λ_i.

If T is small, the distribution of y_{it} may depend on the unknown starting value y_{i0}. In this case, it cannot be assumed that all series $\{y_{i1}, \dots, y_{iT}\}_{i=1,\dots,N}$ are (weakly) stationary. Thus,

two cases arise depending on the treatment of the initial conditions (IC):

(1) **Stationary IC.** If the process has been running for a long time so that all y_{it} are independent of their arbitrary starting values, all T level observations can be used to estimate the forecasting model.

(2) **Non-Stationary IC.** When the process depends on the unobserved starting value y_{i0}, it is convenient to apply a quasi-differencing transformation which removes the dependence on y_{i0} but leaves only $T-1$ usable observations:

$$z_{it} := y_{it} - \rho\, y_{i,t-1} = \lambda_i + \varepsilon_{it}, \qquad i = 1, \ldots, N, \;\; t = 2, \ldots, T \tag{4.2}$$

The resulting series y_{it} or z_{it} (depending on the IC assumption) provide the basis for the analysis of individual effects.

4.2.1 A Decay Test for a Common Initial Shift

Under the assumption of a common shift in the IC across all panel units, the dependence of y_{it} on the initial observation can be tested using a simple decay test. Consider again the dynamic panel model in equation (4.1). Under stationary IC, the initial observation is drawn from the invariant distribution implied by the model,

$$y_{i0} \sim \mathcal{N}\left(\mu_i,\; \frac{\sigma_\varepsilon^2}{1-\rho^2}\right).$$

Under non-stationary IC, the initial observation contains an additional common component whose effect propagates through the AR(1) structure and decays at rate ρ^{t-1}. This implies the auxiliary representation

$$y_{it} = \mu_i + \varsigma\, \rho^{t-1} + u_{it}, \tag{4.3}$$

where ς captures a systematic deviation of the initial condition from its stationary distribution. The test can be implemented by estimating a fixed-effects regression of equation (4.3) using $N \times T$ observations. Inference is based on heteroskedasticity- and autocorrelation-consistent standard errors to account for the serial dependence induced by the dynamic structure. The null hypothesis of stationary initial conditions is

$$H_0 : \varsigma = 0 \quad \text{versus} \quad H_1 : \varsigma \neq 0.$$

Rejection of H_0 provides evidence of a forecast-relevant non-stationarity in the initial condition. Note that the test has power only against a common shift that is nonzero in expectation.

Finite-sample size and power properties are examined in the simulation study in Section 4.4.

4.2.2 The Importance of Precisely Estimated Individual Effects

Even if ρ is known, forecasting accuracy depends on how precisely individual heterogeneity is estimated, especially in short time series. To see this, consider again the process under non-stationary IC. A natural unbiased estimator of the unit-specific effect λ_i is the time average

$$\hat{\lambda}_i = \frac{1}{T-1}\sum_{t=2}^{T} z_{it} = \lambda_i + \bar{\varepsilon}_i, \qquad \bar{\varepsilon}_i := \frac{1}{T-1}\sum_{t=2}^{T} \varepsilon_{it}.$$

Since $\bar{\varepsilon}_i$ averages $T-1$ i.i.d. zero–mean innovations,

$$\mathbb{V}\Big[\hat{\lambda}_i\Big] = \mathbb{V}[\bar{\varepsilon}_i] = \frac{\sigma_\varepsilon^2}{T-1}.$$

Assuming ρ is estimated consistently so that $\hat{\rho} = \rho$, the forecast error can be written as

$$\begin{aligned} e_{i,t+1} &= y_{i,t+1} - \hat{y}_{i,t+1} \\ &= \rho\, y_{it} + \lambda_i + \varepsilon_{i,t+1} - \hat{\rho}\, y_{it} - \hat{\lambda}_i \\ &= \varepsilon_{i,t+1} + \big(\lambda_i - \hat{\lambda}_i\big). \end{aligned}$$

Conditional on λ_i,

$$\mathrm{MSFE}_{i,t+1} = \mathbb{E}\big[e_{i,t+1}^2 \mid \lambda_i\big] = \underbrace{\mathbb{E}\big[\varepsilon_{i,t+1}^2\big]}_{\sigma_\varepsilon^2} + \underbrace{\big(\mathbb{E}[\hat{\lambda}_i] - \lambda_i\big)^2}_{=0} + \underbrace{\mathrm{Var}\Big(\hat{\lambda}_i \mid \lambda_i\Big)}_{\sigma_\varepsilon^2/(T-1)} = \sigma_\varepsilon^2 + \frac{\sigma_\varepsilon^2}{T-1}.$$

The first component, σ_ε^2, is the irreducible innovation uncertainty. The second term $\sigma_\varepsilon^2/(T-1)$ arises from the imprecision in estimating the unit effect λ_i. When $T-1$ is small, this component can be of the same order as the innovation variance itself. This demonstrates that individual-effect estimation error is a key driver of MSFE in short panels, motivating the development of improved estimators in the following sections.

4.2.3 Individual Effect Estimation

This section considers the estimation of individual effects and the construction of optimal forecasts. We assume that the homogeneous parameters ρ, σ_λ^2, and σ_ε^2 are consistently estimated and abstract from additional explanatory variables. The case with explanatory variables is discussed in Section 4.3. Appendix C.1 briefly outlines standard approaches for

the estimation of homogeneous parameters.

Fixed Effects Framework

A fixed-effects (FE) framework treats individual heterogeneity as constant. Thus, individual effects can be estimated by simple plug-in estimates.

Under stationary initial conditions, the level process y_{it} is mean–reverting such that

$$y_{it} = \mu_i + \rho\big(y_{i,t-1} - \mu_i\big) + \varepsilon_{it}, \qquad \mathbb{E}[y_{it}] = \mu_i$$

Hence, the sample average $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ provides an unbiased estimator of μ_i. Substituting this plug–in estimate into the conditional mean yields the FE forecast under stationary IC:

$$\widehat{y}^{FE(S)}_{i,T+1|T} = \bar{y}_i + \rho\big(y_{i,T} - \bar{y}_i\big) = \rho\, y_{i,T} + (1-\rho)\,\bar{y}_i. \tag{4.4}$$

Under non-stationary IC, an analogous procedure based on the ρ–residual series

$$z_{it} = y_{it} - \rho\, y_{i,t-1} = (1-\rho)\mu_i + \varepsilon_{it}, \qquad \mathbb{E}[z_{it}] = (1-\rho)\mu_i$$

is conducted. The sample average $\bar{z}_i = \frac{1}{T-1}\sum_{t=2}^{T} z_{it}$ again provides an unbiased estimator of $(1-\rho)\mu_i$. Substituting this plug–in estimate into the conditional mean gives the non-stationary FE forecast

$$\widehat{y}^{FE(N)}_{i,T+1|T} = \rho\, y_{i,T} + \bar{z}_i. \tag{4.5}$$

Differenced Fixed Effects Framework

As an alternative fixed-effects framework under stationary IC, we may work with the exact finite–sample covariance structure of the first–differenced process rather than the level representation. This eliminates the fixed-effects altogether instead of estimating them. Starting from the model in equation (4.1), we difference both sides to remove the unit–specific mean μ_i, obtaining

$$\Delta y_{it} = \rho\, \Delta y_{i,t-1} + \Delta\varepsilon_{it}.$$

The process in differences can be expanded recursively:

$$\begin{aligned}
\Delta y_{i3} &= \rho\,\Delta y_{i2} + \Delta\varepsilon_{i3},\\
\Delta y_{i4} &= \rho^2\Delta y_{i2} + \rho\,\Delta\varepsilon_{i3} + \Delta\varepsilon_{i4},\\
&\vdots\\
\Delta y_{it} &= \rho^{t-2}\Delta y_{i2} + \sum_{j=0}^{t-3}\rho^j\,\Delta\varepsilon_{i,t-j}.
\end{aligned}$$

Defining the transformed variable $u_{it} := \Delta y_{it} - \rho^{t-2}\,\Delta y_{i2}$, we obtain the analogous recursion

$$\begin{aligned}
u_{i3} &= \Delta\varepsilon_{i3},\\
u_{i4} &= \rho\,\Delta\varepsilon_{i3} + \Delta\varepsilon_{i4},\\
&\vdots\\
u_{it} &= \sum_{j=0}^{t-3}\rho^j\,\Delta\varepsilon_{i,t-j}.
\end{aligned}$$

which shows that $u_{it} = \rho\,u_{i,t-1} + \Delta\varepsilon_{it}$ has an ARMA$(1,1)$ representation. The crucial point is that this recursion starts at $t = 3$; hence the usual stationary ARMA formulas for the autocovariances are only asymptotically valid. To obtain finite–sample exactness, we use a matrix formulation. Let $\varepsilon = (\varepsilon_{i2},\dots,\varepsilon_{iT})'$ and let D be the $(T-2)\times(T-1)$ first–difference matrix. Define the lower–triangular matrix R with ones on its main diagonal and $-\rho$ on the first sub–diagonal. Then the vector of transformed residuals satisfies

$$u = R^{-1}D\varepsilon, \qquad \mathbb{E}[uu'] = \sigma_\varepsilon^2\,R^{-1}DD'(R^{-1})'.$$

The one–step–ahead forecast of the next difference $\Delta y_{i,T+1}$ is obtained as the best linear predictor,

$$\widehat{\Delta y}_{i,T+1|T} = w'\Delta y_{i,2:T},$$

where the optimal weight vector w solves $\Gamma w = \gamma$, with $\Gamma_{ts} = \mathrm{Cov}(\Delta y_{it},\Delta y_{is})$ and $\gamma_t = \mathrm{Cov}(\Delta y_{it},\Delta y_{i,T+1})$ being the autocovariances of the differenced series. Note that γ_t is a model-implied quantity that depends only on ρ and the lag $|T+1-t|$. These autocovariances can be computed either directly from the closed-form expression $\gamma_{|t-s|} = 2\rho^{|t-s|} - \rho^{|t-s+1|} - \rho^{|t-s-1|}$, up to a common factor σ_ε^2 that cancels from the solution, or equivalently via the matrix formulation $\mathbb{E}[uu'] = \sigma_\varepsilon^2\,R^{-1}DD'(R^{-1})'$. Finally, the corresponding level forecast is obtained

by adding the last observed level,

$$\widehat{y}_{i,T+1|T} = y_{iT} + \widehat{\Delta y}_{i,T+1|T}. \tag{4.6}$$

This estimator combines the short–run persistence captured by the differenced autoregression with the exact finite–sample covariance structure implied by the underlying stationary dynamics.

Random Effects Framework

Next, we consider the alternative assumption that λ_i is a random variable, i.e., a random-effects (RE) framework. Under standard assumptions, the resulting forecast is the best linear unbiased predictor (see Goldberger, 1962) .

Consider the reparametrized stationary dynamic panel

$$y_{it} = \rho\, y_{i,t-1} + \lambda_i + \varepsilon_{it},$$

with $\lambda_i \overset{iid}{\sim} (0, \sigma_\lambda^2)$ independent of $\{\varepsilon_{it}\}$. The MSE-optimal one–step-ahead predictor of $y_{i,T+1}$ is the conditional expectation

$$\widehat{y}_{i,T+1|T} = \mathbb{E}\big[y_{i,T+1} \mid y_{i1}, \ldots, y_{iT}\big].$$

Because the level process is the sum of a random intercept and a stationary AR(1), the optimal forecast is a linear projection of $y_{i,T+1}$ on its own history with weights determined by the autocovariance function.

$$\gamma_k := \mathrm{Cov}(y_{it}, y_{i,t-k}) = \sigma_\lambda^2 + \rho^k \frac{\sigma_\varepsilon^2}{1-\rho^2},$$

A detailed derivation of the autocovariance formulas is given in Appendix C.2. Hence, optimal weights $\omega_k = \frac{\gamma_k}{\gamma_0}$ depend not only on the persistence ρ but also on both variance components $(\sigma_\lambda^2, \sigma_\varepsilon^2)$. In practice these covariances can be estimated by OLS or by plug–in estimators of the variance components. When T is large, however, the increasing number of lagged covariances makes the empirical covariance matrix harder to estimate reliably and small errors in its entries can lead to unstable forecast weights. The resulting MSE–optimal forecast under stationary IC and random-effects thus obtains as

$$\widehat{y}^{RE(S)}_{i,T+1|T} = \sum_{t=1}^{T} \omega_t\, y_{i,T+1-t}. \tag{4.7}$$

In the case of non-stationary IC, the one–step MSE–optimal forecast of the residual process obtains as

$$\widehat{z}_{i,T+1|T} = \mathbb{E}\big[z_{i,T+1} \mid z_{i2}, \ldots, z_{iT}\big].$$

Because each process satisfies $z_{it} = \lambda_i + \varepsilon_{it}$ with identical signal λ_i and i.i.d. noise, the best mean–square predictor of λ_i given the transformed history is a reliability–weighted time mean of z_{it}. The explicit weight, derived by linear projection[1], is

$$\kappa = \frac{\sigma_\lambda^2}{\sigma_\lambda^2 + \sigma_\varepsilon^2/(T-1)},$$

so that

$$\mathbb{E}[\lambda_i \mid z_{i2}, \ldots, z_{iT}] = \kappa\, \bar{z}_i, \qquad \bar{z}_i := \frac{1}{T-1}\sum_{t=2}^{T} z_{it}.$$

The coefficient κ is a signal–to–total variance ratio: as $T-1$ increases, the sampling noise $\sigma_\varepsilon^2/(T-1)$ declines and the weight rises; if $\rho \to 1$ so that $\lambda_i = (1-\rho)\mu_i$ vanishes, the weight collapses to zero. Substituting this predictor of λ_i yields the closed–form MSE–optimal forecast under non-stationary IC with random-effects,

$$\widehat{y}_{i,T+1|T}^{RE(N)} = \rho\, y_{iT} + \kappa\, \bar{z}_i. \tag{4.8}$$

4.2.4 Tweedie's Estimator

The *RE* predictor under non-stationarity derived in the previous section shrinks $\bar{z}_i$ toward the cross-sectional mean through the common signal-to-noise ratio κ. The empirical Bayes approach of Liu et al. (2020) generalizes this shrinkage principle by applying Tweedie's formula (Tweedie, 1957; Robbins, 1985) to express the posterior mean of λ_i as a function of the marginal cross-sectional density of $\hat{\lambda}_i$. The key idea is that the residuals z_{it} provides a natural sufficient statistic for λ_i. Their time-series average

$$\hat{\lambda}_i \;=\; \bar{z}_i \;=\; \frac{1}{T-1}\sum_{t=2}^{T} z_{it}, \qquad \hat{\lambda}_i \mid \lambda_i \;\sim\; \mathcal{N}\!\left(\lambda_i,\, \frac{\sigma_\varepsilon^2}{T-1}\right)$$

is an unbiased estimator of λ_i, but noisy when T is small. Tweedie's formula corrects for this noise by expressing the posterior mean of λ_i as a function of the marginal cross-sectional density of $\hat{\lambda}_i$,

$$\mathbb{E}\Big[\lambda_i \mid \hat{\lambda}_i\Big] \;=\; \hat{\lambda}_i \;+\; \frac{\sigma_\varepsilon^2}{T-1}\,\frac{\partial}{\partial \hat{\lambda}_i}\ln p\Big(\hat{\lambda}_i\Big),$$

[1] A detailed derivation of this conditional mean is given in Appendix C.3.

where $p(\hat{\lambda}_i)$ is estimated from the observed $\hat{\lambda}_i$ across units. Therefore, the prior on λ_i enters only through $p(\hat{\lambda}_i)$ and never needs to be specified explicitly. The correction term pulls extreme estimates toward the center of the cross-sectional distribution. As discussed by Efron (2011), large estimates often appear extreme for two reasons: the corresponding effects are indeed sizable, but random disturbances have additionally increased their observed values beyond the underlying signal. Thus, the average of the largest estimates is biased upward; a classic form of the winner's curse, also known as selection bias. The one-step-ahead forecast then follows as

$$\hat{y}_{i,T+1|T}^{TW(N)} \;=\; \mathbb{E}\Big[\lambda_i \mid \hat{\lambda}_i\Big] \;+\; \rho\, y_{iT}. \tag{4.9}$$

In the parametric implementation, $p(\hat{\lambda}_i)$ is approximated by a finite mixture of K normal distributions,

$$\hat{p}_{\text{mix}}(\hat{\lambda}_i) \;=\; \sum_{k=1}^{K} \omega_k \,\mathcal{N}\Big(\hat{\lambda}_i \mid \mu_k,\, \sigma_k^2\Big)\,, \qquad \omega_k \geqslant 0, \quad \sum_{k=1}^{K} \omega_k = 1,$$

where the parameters are estimated by maximum likelihood and K is selected via pseudo-out-of-sample forecast errors.

In the nonparametric kernel implementation, Liu et al. (2020) propose estimating $p(\hat{\lambda}_i)$ by a leave-one-out kernel density estimator, where the bandwidth is selected via pseudo-out-of-sample forecast errors. The authors prove that as $N \to \infty$ with T fixed, the estimator "achieves the same compound risk as the oracle predictor" that knows the true distribution of λ_i, a property they call ratio-optimality.

4.3 Dynamic Models with Explanatory Variables

In this section, the baseline model in (4.1) is extended to include additional explanatory variables. A distinction is made according to whether they are time-invariant or time-varying.

4.3.1 Time-invariant Explanatory Variables

We begin with the dynamic panel model for the case of time-invariant explanatory variables:

$$y_{it} = \rho y_{i,t-1} + x_i'\gamma + \eta_i + \varepsilon_{it}, \tag{4.10}$$

where the individual effect satisfies $\lambda_i = x_i'\gamma + \eta_i$. The implications for forecasting differ between fixed- and random-effects approaches.

Fixed Effects

Under a fixed-effects specification, see equations (4.4), (4.5) and (4.6), individual heterogeneity is treated as an unrestricted parameter. Since x_i is time-invariant, it is perfectly collinear with the unit-specific intercept and therefore absorbed into the estimated effect. Thus, the inclusion of x_i does not affect the forecasting rule.

Random Effects

Under a random-effects specification, see equations (4.7) and (4.8), the decomposition $\lambda_i = x_i'\gamma + \eta_i$ implies that part of the persistent heterogeneity is explained by observables. Forecasting is then based on

$$\mathbb{E}[\lambda_i \mid y_{i1}, \ldots, y_{iT}] = x_i'\gamma + E[\eta_i \mid y_{i1}, \ldots, y_{iT}]$$

so that shrinkage applies only to the residual component η_i. Under stationary IC, the optimal predictor follows from the RE formulas in Section 4.2.3, with the covariance structure determined by σ_η^2 instead of σ_λ^2. Under non-stationary IC, the quasi-differenced representation implies $z_{it} = x_i'\gamma + \eta_i + \varepsilon_{it}$, and the MSE–optimal RE forecast becomes

$$\widehat{y}_{i,T+1|T} = \rho y_{iT} + x_i'\gamma + \kappa_\eta \bar{\eta}_i,$$

where $\bar{\eta}_i = \bar{z}_i - x_i'\hat{\gamma}$ and $\kappa_\eta = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/(T-1)}$. Hence, unlike the FE case, time-invariant regressors affect forecasts by reducing the variance of the latent component subject to shrinkage.

4.3.2 Time-varying Explanatory Variables

Now consider the dynamic panel model for the case with time-varying explanatory variables:

$$y_{it} = \rho y_{i,t-1} + x_{it}'\beta + \lambda_i + \varepsilon_{it}. \tag{4.11}$$

Under stationary IC, subtracting the contemporaneous regressor component does not suffice. Define $\tilde{y}_{it} = y_{it} - x_{it}'\beta$ and consider the model that subtracts only $x_{it}'\beta$:

$$\begin{aligned}
\tilde{y}_{it} &= \rho y_{i,t-1} + \lambda_i + \varepsilon_{it} \\
&= \rho(\tilde{y}_{i,t-1} + x_{i,t-1}'\beta) + \lambda_i + \varepsilon_{it} \\
&= \rho\tilde{y}_{i,t-1} + \rho x_{i,t-1}'\beta + \lambda_i + \varepsilon_{it}.
\end{aligned}$$

The residuals would therefore still depend on lagged regressors and not satisfy the covariance structure of a pure AR(1) model with random intercept. Moment-based weights derived under stationary IC are consequently misspecified in the presence of time-varying regressors.

The quasi-differencing transformation, by contrast, removes both the dynamic and contemporaneous regressor components simultaneously. The resulting full residual series thus coincides with the specification analyzed in Section 4.2.3:

$$\begin{aligned} z_{it} &= y_{it} - \rho y_{i,t-1} - x_{it}'\beta \\ &= \lambda_i + \varepsilon_{it}, \end{aligned}$$

However, the inclusion of time-varying regressors raises the question of how λ_i relates to $\{x_{it}\}_{t=1}^{T}$. If λ_i is correlated with the regressors, a pure random-effects specification is inconsistent, while a fixed-effects approach remains valid but inefficient. Following Mundlak (1978) and Chamberlain (1982), this trade-off can be resolved by decomposing

$$\lambda_i = c_i'\gamma + \eta_i,$$

with $\eta_i \overset{iid}{\sim} (0, \sigma_\eta^2)$, $\mathbb{E}(\eta_i c_i) = 0$ and c_i collects time-invariant functions of the regressor history. The resulting correlated-effects (CE) approach is structurally identical to the one in the time-invariant case, with the distinction that c_i is now constructed from the observed regressors. The Mundlak specification sets $c_i = \bar{x}_i$, while the Chamberlain specification uses the full history $c_i = (x_{i1}', \ldots, x_{iT}')'$, allowing the informational content for λ_i to vary over time.

Fixed Effects

Under a FE specification, any time-invariant correlated component $c_i'\gamma$ is again absorbed by the unit-specific intercept. Thus, the CE decomposition is irrelevant under fixed effects.

Random Effects

Under a RE specification, the CE decomposition $\lambda_i = c_i'\gamma + \eta_i$ allows part of the persistent heterogeneity to be explained by observables, so that shrinkage applies only to the residual component η_i. The quasi-differenced series becomes $z_{it} = c_i'\gamma + \eta_i + \varepsilon_{it}$ and the MSE-optimal one-step-ahead forecast is

$$\widehat{y}_{i,T+1|T}^{CE} = \rho y_{iT} + x_{i,T+1}'\beta + c_i'\gamma + \kappa_\eta \bar{\eta}_i, \tag{4.12}$$

where $\bar{\eta}_i = \bar{z}_i - c_i'\hat{\gamma}$ and

$$\kappa_\eta = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/(T-1)}.$$

The CE predictor benefits from c_i in two ways. First, since $\sigma_\eta^2 \leqslant \sigma_\lambda^2$, the shrinkage factor κ_η is smaller than the corresponding κ in the baseline RE predictor, reducing the variance of the latent component subject to shrinkage. Second, $c_i'\gamma$ enters directly into the conditional mean of $y_{i,T+1}$, thereby enhancing predictive accuracy whenever λ_i is correlated with the regressors.

4.4 Simulation

To study the finite-sample forecasting performance of the forecasting methods considered in this chapter, we conduct a series of Monte Carlo experiments based on the dynamic panel model

$$y_{it} = \lambda_i + \rho\, y_{i,t-1} + x_{it}'\beta + \varepsilon_{it}, \qquad i = 1, \ldots, N, \;\; t = 1, \ldots, T.$$

ε_{it} is drawn from a standard normal distribution, e.g., $\varepsilon_{it} \sim \mathcal{N}(0,1)$.[2] Individual effects (distributed as $\mathcal{N}(0, \sigma_\lambda^2)$) and innovations are always independent across i and t.

We consider data-generating environments with and without covariates, and with stationary or non-stationary initial conditions. In the stationary setup, the process is initialized from its unconditional distribution and we discard a burn-in period of 100 observations before retaining the final T periods used for estimation and forecasting. In the non-stationary setup, we set $y_{i0} = 10$ and simulate the T observations without any burn-in.

In each Monte Carlo experiment we vary one of the key parameters $(\sigma_\lambda^2, \rho, T, N)$ over a five-point grid, holding the remaining parameters fixed at their baseline values $(\sigma_\lambda^2, \rho, T, N) = (0.5,\ 0.8,\ 10,\ 200)$. The full grids used in the simulations are:

$$\begin{aligned}
\sigma_\lambda^2 &\in \{0.05,\ 0.25,\ \mathbf{0.5},\ 1,\ 2\},\\
\rho &\in \{-0.2,\ 0.5,\ \mathbf{0.8},\ 0.95,\ 0.98\},\\
T &\in \{5,\ \mathbf{10},\ 25,\ 50,\ 200\},\\
N &\in \{20,\ 100,\ \mathbf{200},\ 500,\ 1000\}.
\end{aligned}$$

The values in each grid are meant to cover a broad range of empirically relevant situations. For the individual-effect variance σ_λ^2, the grid spans settings with almost no (0.05) to very

[2] Within a robustness check, also non-Gaussian errors from a centered chi-squared distribution ($\varepsilon_{it} \sim \chi_1^2 - 1$) are considered. The results do not differ substantially from those obtained under normally distributed errors.

large unobserved heterogeneity (1 and 2). Since the unconditional variance of y_{it} under a standard normal error distribution is $\text{Var}(y_{it}) = \sigma_\lambda^2 + \frac{1}{1-\rho^2}$, changing σ_λ^2 from 0.05 to 2 shifts the contribution of heterogeneity from 1.8% to 41.9% of the total variance under the baseline $\rho = 0.8$. The grid for ρ includes one case of negative dependence and four cases with positive autocorrelation that move from moderate persistence to near unit root behavior. The time and the cross-sectional grids let us investigate very short to moderately long time series as well as different dimensions of the cross section.

For each parameter configuration, we generate $R = 1{,}000$ independent data sets. All forecasting exercises are applied to the estimation sample of length T, and the resulting one-step-ahead forecast errors are collected. Forecast performance is evaluated by the mean of the mean squared forecast error.

4.4.1 Testing for Non-Stationary Initial Conditions

We first investigate the finite-sample size and power properties of the proposed decay test of Section 4.2.1. To isolate the properties of the test, we treat ρ as known. Under stationary initial conditions, the initial observation is drawn from the invariant distribution,

$$y_{i0} \sim \mathcal{N}\left(\mu_i,\ \frac{\sigma_\varepsilon^2}{1-\rho^2}\right).$$

Under non-stationary initial conditions, we study the effect of a mean shift measured in units of the stationary standard deviation of the process

$$y_{i0} \sim \mathcal{N}\left(\mu_i + a\sqrt{\tfrac{\sigma_\varepsilon^2}{1-\rho^2}},\ \frac{\sigma_\varepsilon^2}{1-\rho^2}\right),$$

where the parameter $a \geqslant 0$ controls the magnitude of the deviation from stationarity. We consider a grid of values $a \in [0, 0.5]$, where $a = 0$ corresponds to stationary initial conditions and $a > 0$ indicates deviations from stationarity. For each value of a, we compute the empirical rejection probability at the 5% level. Rejection frequencies at $a = 0$ are interpreted as size, while rejection frequencies for $a > 0$ represent power. The following figure depicts the results based on $1,000$ iterations.

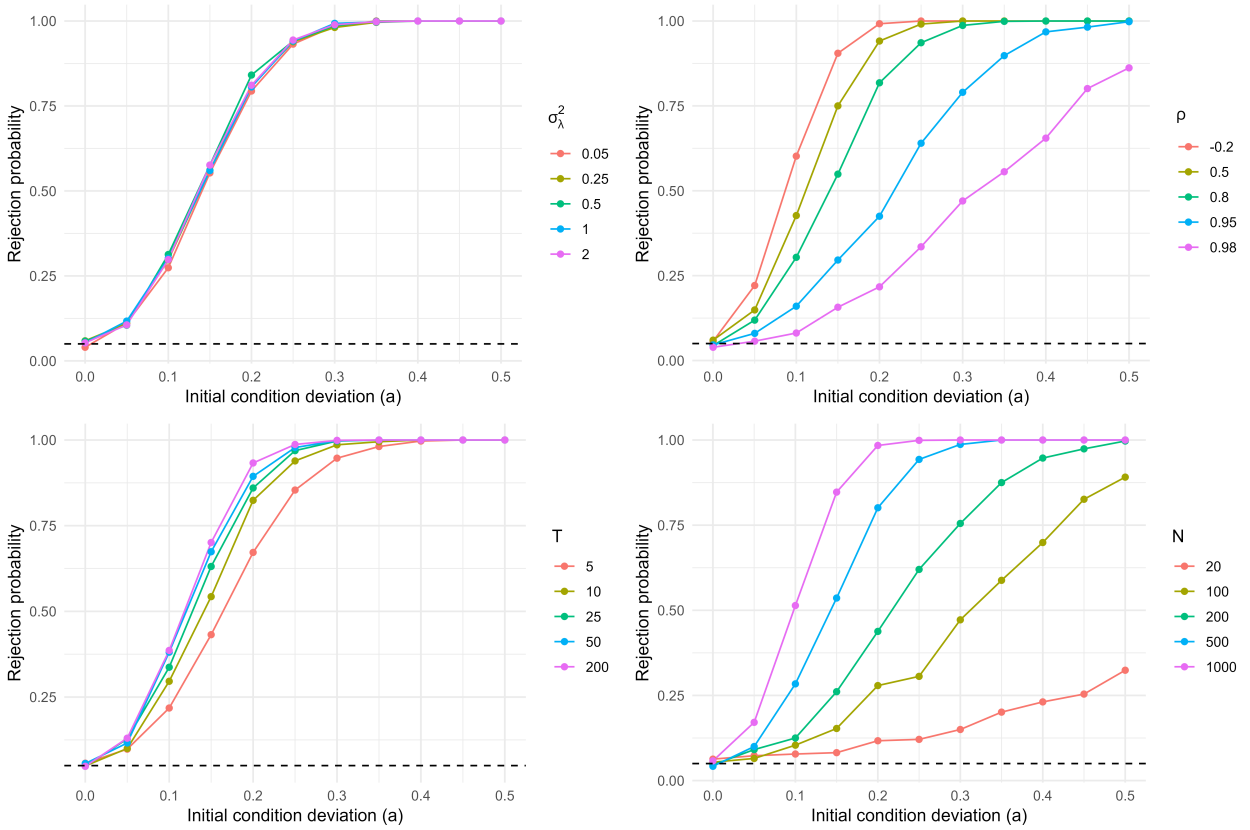

Figure 4.1: Empirical power curves of the decay test. The dashed lines indicate the 5%-rejection level.

The simulation results reveal several intuitive patterns. First, size and power of the test are largely unaffected by the magnitude of the individual variance σ^2_λ. This reflects that the fixed-effects transformation removes time-invariant heterogeneity across units. Second, the power of the test declines as the persistence parameter ρ increases. When ρ is close to one, the decay term ρ^{t-1} changes only slowly over time. As a result, the additional component in the initial condition becomes harder to distinguish from the fixed-effects, which reduces the ability of the test to detect deviations from stationarity. Finally, the test is more sensitive to the cross-sectional dimension N than to the time dimension T. Since the decay regressor varies only over time, identification mainly relies on estimating cross-sectional averages precisely. Increasing N therefore reduces sampling variability. In contrast, increasing T adds relatively little information because the signal decays geometrically over time.

An important limitation of the decay test is that it has no power against cross-sectionally heterogeneous initial shifts with mean zero. Detecting this type of non-stationarity would require exploiting second-moment implications, which is more demanding in finite samples because identification would then rely primarily on variation over the typically small time dimension rather than on the larger cross-sectional dimension.

4.4.2 DGP without Covariates

This section studies the performance of the proposed predictors for data-generating processes without explanatory variables. In a first step the parameter ρ is estimated via GMM of the differenced model equation as proposed by Arellano and Bond (1991).

Stationary Initial Conditions

The first Monte Carlo experiment compares the forecast accuracy across the different parameter grids under stationary IC. Table 4.1 reports the unconditional MSFE across the four parameter grids for the DGP without covariates and stationary initial conditions. Overall, the RE predictor under stationary initial conditions achieves the lowest MSFE in 14 out of 20 cases. In some settings such as near–unit-root behavior ($\rho \geqslant 0.95$), short time series ($T = 5$), and small cross sections ($N = 20$) methods designed for non-stationary IC can (slightly) outperform the stationary RE predictor. The $FE_D(S)$ predictor performs slightly worse, ranking second in 12 out of 20 cases, but exhibits weaknesses in the same situations as the RE predictor. By contrast, the RE predictor for non-stationary IC serves as a robust alternative, performing particularly well precisely when its stationary counterpart struggles. The two predictors based on Tweedie's empirical Bayes approach are optimal in certain edge cases ($\sigma_\lambda^2 = 0.05$, $\rho = -0.2$, and $N = 20$), but are not able to systematically outperform the alternative predictors.

Table 4.1: Average MSFE for stationary IC with $\mathcal{N}(0,1)$ errors. The DGP does not include covariates. The MSFE of the best-performing model is shown in bold.

Parameter	FE(S)	FE(N)	$\mathbf{FE_D(S)}$	RE(S)	RE(N)	$\mathbf{TW_P(N)}$	$\mathbf{TW_K(N)}$
$\sigma_\lambda^2 = 0.05$	1.0784	1.1206	1.0721	1.0413	1.045	**1.0392**	1.0644
$\sigma_\lambda^2 = 0.25$	1.0544	1.0931	1.045	**1.0381**	1.063	1.0576	1.0619
$\sigma_\lambda^2 = 0.5$	1.0861	1.1214	1.0777	**1.0724**	1.1009	1.2678	1.154
$\sigma_\lambda^2 = 1$	1.0675	1.0945	1.0552	**1.0531**	1.0806	1.4217	1.2375
$\sigma_\lambda^2 = 2$	1.1534	1.1579	1.1365	**1.1353**	1.1505	1.8389	1.6877
$\rho = -0.2$	1.1089	1.1157	1.1084	1.0909	1.0959	**1.0905**	1.0964
$\rho = 0.5$	1.0948	1.1223	1.0886	**1.0781**	1.1028	1.0784	1.0926
$\rho = 0.8$	1.0718	1.1176	1.063	**1.0575**	1.0924	1.2512	1.1374
$\rho = 0.95$	1.5814	1.4602	1.5574	1.554	**1.4562**	1.9434	3.7074
$\rho = 0.98$	2.0919	1.8833	2.069	2.0689	**1.8815**	1.97	38.0697
$T = 5$	**1.1633**	1.2623	1.1765	1.2301	1.1967	2.2131	2.4241
$T = 10$	1.0871	1.1184	1.0777	**1.074**	1.0997	1.2826	1.1583
$T = 25$	1.0402	1.0444	1.0365	**1.0356**	1.0426	1.0737	1.0453
$T = 50$	1.029	1.0311	1.0264	**1.0262**	1.0307	1.0356	1.0273
$T = 200$	0.9976	**0.9974**	**0.9974**	**0.9974**	**0.9974**	0.9976	**0.9974**
$N = 20$	1.3428	1.3167	1.3288	1.3289	1.3091	1.3121	**1.172**
$N = 100$	1.0983	1.1193	1.0862	**1.0811**	1.0988	1.2783	1.1445
$N = 200$	1.0791	1.1135	1.0703	**1.0651**	1.0939	1.2838	1.1602
$N = 500$	1.0564	1.1015	1.0493	**1.0453**	1.0824	1.2739	1.1631
$N = 1000$	1.0731	1.1204	1.0672	**1.0622**	1.1014	1.3162	1.2112

Abbreviations:
$FE(S)$ = Fixed effects estimator for stationary IC (4.4),
$FE(N)$ = Fixed effects estimator for non-stationary IC (4.5),
$FE_D(S)$ = Stationary fixed effects estimator based on first differences (4.6),
$RE(S)$ = Random effects estimator for stationary IC (4.7),
$RE(N)$ = Random effects estimator for non-stationary IC (4.8),
$TW_P(N)$ = Parametric Tweedie estimator for non-stationary IC (4.9),
$TW_K(N)$ = Kernel Tweedie estimator for non-stationary IC (4.9).

The first five rows of Table 4.1 vary the variance of the individual effects. As σ_λ^2 increases, MSFE rises for all methods, reflecting the increasing difficulty of recovering unit-specific components in short panels as higher cross-sectional dispersion mechanically reduces the informativeness of pooled moments. Methods designed for stationary initial conditions dominate their non-stationary counterparts uniformly across the grid, confirming that avoiding differencing preserves valuable time-series information when T is small. Among these, the RE predictor under stationary IC performs best overall and remains stable as heterogeneity increases. The FE_D forecast also performs well at higher levels of σ_λ^2, indicating that it can be advantageous to avoid the explicit estimation of λ_i when σ_λ^2 is large. The parametric

empirical Bayes estimator is particularly sensitive to the degree of heterogeneity. When σ_λ^2 is small the cross-sectional distribution of $\hat{\lambda}_i$ is narrow and well-approximated by a Gaussian mixture, but as dispersion increases the density estimate becomes unreliable in the tails where the Tweedie correction is largest.

Rows 6-10 consider different autoregressive coefficients. As ρ approaches unity, forecast accuracy deteriorates for all methods. This behavior is expected: as the second moment scales with $(1-\rho^2)^{-1}$, small estimation errors in ρ can lead to large distortions when persistence is high. In addition, for large ρ the variance induced by strong autocorrelation dominates cross-sectional heterogeneity, which further limits the gains from pooling. For moderate persistence ($\rho \leqslant 0.8$), stationary panel methods dominate their non-stationary counterparts. For very high persistence ($\rho \geqslant 0.95$), the ranking changes and methods designed for non-stationary IC outperform their stationary counterparts. Although the first observation is discarded, this yields more stable estimates by effectively reducing persistence. The Tweedie estimators become unstable as ρ approaches unity. In near-unit-root settings, finite-sample errors in $\hat{\rho}$ contaminate the sufficient statistic $\bar{z}_i$, causing the kernel density derivative to become numerically unreliable.

Rows 11-20 vary T and N. Forecast accuracy improves for all methods as the length of the time series increases, reflecting the additional information available for estimating both dynamic and unit-specific components. As T becomes large, performance differences across alternative predictors vanish. For $T = 50$, the MSFE spread between the best and worst panel methods is below one percent, and for $T = 200$ it becomes negligible. In this region, cross-sectional shrinkage plays only a minor role and most methods effectively approximate the same forecasting rule. Increasing N also improves forecast accuracy for the majority of methods. The two estimators based on Tweedie's formula show less sensitivity. Most of the improvement occurs as N increases from very small to moderate values ($N \geqslant 100$); beyond that, additional gains are fairly limited. Across the grid, the stationary RE predictor again delivers the lowest MSFE. For $N = 20$, methods designed for non-stationary IC perform relatively well, likely because they rely less on cross-sectional information as a shrinkage target. In particular, the kernel-based Tweedie estimator exhibits strong shrinkage in this setting.

Non-Stationary Initial Conditions

Analogously to the previous table, Table 4.2 depicts the average MSFE resulting from the DGP with non-stationary IC. An important asymmetry emerges. Applying methods designed for non-stationary IC to data with stationary IC is inefficient but not incorrect. By contrast,

applying methods designed for stationary IC to non-stationary data involves a bias and leads to substantial losses in forecast accuracy. These methods rely on information on the level y_{it} and implicitly condition on moments of the invariant distribution, which are not reliable when the process has not yet converged to its stationary regime. Consistent with this asymmetry, the *RE* predictor for non-stationary IC delivers the lowest MSFE in 16 out of 20 cases, followed by the parametric and kernel-based Tweedie predictors, which are always among the top two models. Conversely, the $FE_D(S)$ approach performs poorly throughout. This outcome is expected, since differencing transmits the non-stationary initialization into the regressors, leading to biased forecasts under non-stationary IC.

Table 4.2: Average MSFE for non-stationary IC with $\mathcal{N}(0,1)$ errors. The DGP does not include covariates. The MSFE of the best-performing model is shown in bold. For further notes, see Table 4.1.

Parameter	**FE(S)**	**FE(N)**	**FE$_D$(S)**	**RE(S)**	**RE(N)**	**TW$_P$(N)**	**TW$_K$(N)**
$\sigma^2_\lambda = 0.05$	1.8337	1.1029	2.0326	1.2456	**1.026**	1.0262	1.0512
$\sigma^2_\lambda = 0.25$	1.9069	1.1283	2.1128	1.7921	**1.0941**	1.0952	1.1054
$\sigma^2_\lambda = 0.5$	1.9433	1.1153	2.1643	1.988	**1.092**	1.0926	1.0983
$\sigma^2_\lambda = 1$	2.0349	1.1062	2.2807	2.1977	**1.095**	1.0952	1.0983
$\sigma^2_\lambda = 2$	2.2925	1.1278	2.5932	2.5638	1.1225	**1.121**	1.1229
$\rho = -0.2$	2.0809	1.099	1.7664	1.5454	**1.082**	1.0826	1.0872
$\rho = 0.5$	2.0416	1.0976	2.5986	2.2006	**1.0792**	1.0814	1.0858
$\rho = 0.8$	1.9587	1.1061	2.181	2.0007	**1.0879**	1.09	1.0932
$\rho = 0.95$	1.5316	1.105	1.5426	1.5338	1.088	**1.083**	1.0894
$\rho = 0.98$	1.7654	1.1643	1.7564	1.7719	1.174	**1.0954**	1.1013
$T = 5$	3.0636	1.2486	3.1827	2.7157	**1.1638**	1.1688	1.1952
$T = 10$	1.9551	1.1135	2.1794	1.9981	**1.0915**	1.0921	1.0993
$T = 25$	1.1793	1.0265	1.324	1.2996	**1.0218**	1.0233	1.0247
$T = 50$	1.0462	1.0064	1.1061	1.1036	**1.0056**	1.0061	1.0058
$T = 200$	0.9918	0.989	0.9972	0.9974	0.9889	0.9889	**0.9888**
$N = 20$	2.0091	1.0982	2.2442	2.0454	**1.0659**	1.0715	1.0965
$N = 100$	1.9879	1.1319	2.211	2.0335	**1.1123**	1.1134	1.1195
$N = 200$	1.9542	1.0929	2.1792	2.0021	**1.0745**	1.0756	1.0787
$N = 500$	1.9411	1.1172	2.16	1.9892	**1.0954**	1.0955	1.1001
$N = 1000$	1.9369	1.1092	2.1547	1.983	**1.0887**	1.0889	1.091

As σ^2_λ increases, MSFEs rise again for all methods, with especially pronounced deterioration for procedures designed for stationary IC. Methods designed for non-stationary IC now uniformly outperform their stationary counterparts. This difference is driven primarily by the poor performance of stationary-IC methods under non-stationary initialization, rather than by substantial improvements of non-stationary-IC methods relative to the stationary DGP. Overall, the *RE* predictor for non-stationary IC performs best and remains stable as

heterogeneity increases. The Tweedie-based predictors perform substantially better than in the stationary case and are consistently among the top-performing methods.

As in the stationary case, forecast accuracy deteriorates as ρ approaches unity, but the deterioration is less severe. Across the grid, methods designed for non-stationary IC outperform their stationary counterparts. For moderate persistence ($\rho \leqslant 0.8$), the *RE* predictor for non-stationary IC performs best overall. As persistence becomes very high, this predictor deteriorates relative to the Tweedie-based predictors, which remain stable and dominate for $\rho = 0.95$ and $\rho = 0.98$. This contrasts with the stationary DGP, where the Tweedie estimators became highly unstable.

Similar to the case with stationary initialization, forecast accuracy improves for all methods as T increases. In contrast to the stationary case, performance differences across panel estimators persist for larger values of T, reflecting the fact that additional time is required for the influence of the non-stationary starting value to decay. Across the grid, methods designed for non-stationary IC dominate, with the *RE* predictor performing best overall. The last rows vary N. The qualitative pattern mirrors that under stationary initialization, but with more pronounced losses from applying stationary-IC methods to non-stationary data. Across the grid, the *RE* predictor for non-stationary IC performs best.

4.4.3 DGP with Covariates

We extend the baseline data-generating process with stationary IC by including $K = 3$ observed covariates collected in the vector $x_{it} = (x_{1it}, x_{2it}, x_{3it})'$ that are predetermined and whose components follow independent stationary AR(1) processes with persistence parameter $\phi_x = 0.8$ and variance $\sigma_x^2 = 0.1$. The initial observations are drawn from the corresponding stationary distribution. Individual effects λ_i are first generated independently of the regressors, before we study the role of correlated effects (CE) within a Mundlak- and a Chamberlain-inspired approach.

Independent Individual Effects

First, we consider the case where individual effects are independent of the observed covariates.[3] Prior to forecasting, the common parameters $(\hat{\rho}, \hat{\beta})$ are again estimated via GMM. Especially in samples with small N, forecasting is therefore based on a process with composite error term $\widetilde{\varepsilon}_{it}$ that absorbs the estimation error arising from $(\hat{\rho} - \rho)$ and $(\hat{\beta} - \beta)$. This contamination increases noise in the estimation of individual effects and second moments,

[3]The results reported in this subsection are based on a preliminary Monte Carlo study with a limited number of 100 replications and are intended to be illustrative.

thereby undermining the calibration of the shrinkage factor κ. An additional complication arises for predictors derived under stationary initial conditions. As outlined in Section 4.3, only models based on non-stationary IC are correctly specified, since in the stationary case the lagged dependent variable still embeds past realizations of the regressors. Consistent with this argument, models based on non-stationary IC dominate their stationary counterparts for all parameter values, with average MSFE reductions between 15.9% (RE) and 20.5% (FE). Moreover, conditioning on x_{it} with estimated common parameters and the correctly specified non-stationary IC improves forecast performance under the DGP with covariates by an average of 17.8% (RE) and 22.9% (FE) in terms of MSFE compared to non-stationary models without covariates.

Correlated Effects

We now study the role of correlated effects (CE) in more detail. Unlike the preliminary results in the previous subsection, the following analysis is based on a large-scale Monte Carlo study with $R = 1{,}000$ replications per parameter configuration, following the design described in Section 4.4. To isolate this mechanism, the simulation conditions on the true common parameters (ρ, β) so that differences in forecast performance arise solely from how the individual effects are modeled. Correlated effects are introduced by augmenting the regressors with components proportional to the individual effect λ_i. Starting from the baseline AR(1) process x_{itk}, the observed regressors are generated as

$$x_{itk}^{CE} = x_{itk} + c_{tk}\lambda_i,$$

where c_{tk} denotes the loading of λ_i on regressor k at time t. In the Mundlak design the loading is time-invariant, $c_{tk} = c_k$, with $(c_1, c_2, c_3) = (1, -1, 0)$. In the Chamberlain design the loading varies over time. Specifically, we use a geometrically decaying pattern $c_t = 0.8^{T-t} - \bar{c}$ and set $(c_{t1}, c_{t2}, c_{t3}) = (c_t, -c_t, 0)$. This specification preserves the symmetric loading structure of the Mundlak design while allowing the information about λ_i contained in the regressors to decline with the distance from the forecast period.

Forecast performance is evaluated for four predictors. First, we consider the FE predictor and the RE shrinkage predictor that ignore correlated effects, corresponding to the baseline models discussed in Section 4.2.3. Second, we consider two correlated-effects extensions of the RE predictor based on the introduced Mundlak and Chamberlain approaches.[4] In both cases

[4] In the fixed-effects model any time-invariant correlated component is absorbed by the unit-specific intercept and therefore cannot affect the FE forecast; see Section 4.3. Further, we exclude the predictor based on Tweedie because it was never competitive but increased computation time substantially.

the individual component is approximated by regressing the transformed residual average $\bar{z}_i$ on functions of the regressors, namely the individual means (Mundlak) or the full regressor history (Chamberlain). To mitigate overfitting, coefficients are estimated in two steps. We first estimate the full model and retain coefficients significant at the 5% level, then re-estimate the model using only the significant coefficients. Table 4.3 shows the results.

Table 4.3: Average MSFE for stationary IC with 3 covariates under different CE specifications

	No CE in DGP				Mundlak CE in DGP				Chamberlain CE in DGP			
Parameter	FE	RE	RE^M	RE^C	FE	RE	RE^M	RE^C	FE	RE	RE^M	RE^C
$\sigma^2_\lambda = 0.05$	1.113	**1.0354**	1.0359	1.0369	1.1124	1.0364	**1.0278**	1.0349	1.1131	**1.0367**	1.0371	1.038
$\sigma^2_\lambda = 0.25$	1.1165	**1.0814**	1.0817	1.0821	1.109	1.0753	**1.039**	1.0595	1.1064	1.073	1.0693	**1.0644**
$\sigma^2_\lambda = 0.50$	1.1142	**1.0935**	1.0937	1.0939	1.1204	1.0991	**1.0499**	1.0743	1.1107	1.0912	1.0861	**1.077**
$\sigma^2_\lambda = 1.00$	1.111	**1.0994**	1.0995	1.0997	1.115	1.104	**1.0476**	1.0732	1.1101	1.0994	1.0925	**1.0808**
$\sigma^2_\lambda = 2.00$	1.1138	**1.1081**	1.1082	1.1082	1.1123	1.107	**1.0484**	1.0771	1.1088	1.1033	1.0957	**1.0826**
$\rho = -0.20$	1.1099	**1.0903**	1.0905	1.0908	1.113	1.092	**1.0434**	1.0676	1.1147	1.0946	1.0884	**1.0791**
$\rho = 0.50$	1.1067	**1.0877**	1.0879	1.0882	1.1103	1.0905	**1.041**	1.0662	1.112	1.0912	1.0854	**1.0766**
$\rho = 0.80$	1.1104	**1.0918**	1.0919	1.0922	1.1125	1.092	**1.0432**	1.0674	1.1115	1.0916	1.0854	**1.0752**
$\rho = 0.95$	1.116	**1.0959**	1.0961	1.0963	1.1148	1.0942	**1.046**	1.0693	1.1123	1.0912	1.0853	**1.0762**
$\rho = 0.98$	1.1049	**1.0843**	1.0845	1.0847	1.1133	1.0934	**1.0432**	1.0689	1.1091	1.0885	1.0822	**1.0726**
$T = 5$	1.2526	**1.1683**	1.1695	1.1692	1.2527	1.1722	**1.0703**	1.119	1.2536	1.1712	1.1393	**1.1294**
$T = 10$	1.1142	**1.0947**	1.0949	1.0952	1.1151	1.0949	**1.0451**	1.0698	1.1105	1.0899	1.0843	**1.0747**
$T = 25$	1.0431	**1.0397**	1.0397	1.0399	1.0419	1.0384	**1.0206**	1.0317	1.043	1.0399	1.0396	**1.0368**
$T = 50$	1.0167	**1.0158**	1.0158	1.0158	1.0186	1.0177	**1.0094**	1.0151	1.0215	1.0206	1.0206	**1.0201**
$T = 200$	1.0057	**1.0056**	1.0056		1.0075	1.0075	**1.0056**		**1.0025**	1.0026	1.0026	
$N = 20$	1.0993	**1.0813**	1.0836		1.1184	1.1012	**1.073**		1.0826	**1.0669**	1.0696	
$N = 100$	1.1201	**1.1003**	1.1008	1.1014	1.113	1.0929	**1.0451**	1.0782	1.1088	1.0896	1.0843	**1.0795**
$N = 200$	1.1112	**1.0925**	1.0927	1.093	1.1148	1.0951	**1.0457**	1.0719	1.1129	1.0939	1.0884	**1.0778**
$N = 500$	1.1138	**1.0937**	1.0938	1.0938	1.1113	1.0905	**1.0415**	1.0521	1.1117	1.0918	1.0856	**1.0746**
$N = 1000$	1.1112	**1.0909**	1.0909	1.091	1.1098	1.0903	**1.0407**	1.044	1.1102	1.0898	1.0835	**1.0725**

The DGP does include covariates. The models are estimated with and without correlated effects (CE). The MSFE of the best-performing model is shown in bold.

Abbreviations:
FE = Fixed effects estimator for non-stationary IC without CE,
RE = Random effects estimator for non-stationary IC without CE,
RE^M = Random effects estimator for non-stationary IC with Mundlak-CE,
RE^C = Random effects estimator for non-stationary IC with Chamberlain-CE.

The following results can be recorded. First, as expected, predictors that correctly reflect the DGP perform best. Consistent with the simulations without explanatory variables, the RE predictor generally dominates the FE predictor, with both approaches converging when the time dimension becomes very large ($T = 200$). Second, the relative performance of the CE approaches depends on the amount of heterogeneity in the data. When heterogeneity is very small, the regressors contain little information about the individual effects, so that the Mundlak and baseline RE predictors perform better than the Chamberlain specification. Moreover, the Chamberlain approach relies on an unregularized projection involving the full regressor history and therefore cannot be estimated in configurations where the number of regressors exceeds the cross-sectional dimension ($T = 200$ and $N = 20$). Third, an important question is how CE predictors perform when the regressors are in fact orthogonal to the individual effects. In this case the Mundlak approach proves to be virtually risk-free: even under the DGP without correlated effects it performs only marginally worse than the correctly specified RE model (at most 0.2% for $N = 20$ and on average 0.03% across all parameter combinations). The Chamberlain approach exhibits similarly small deviations. However, when correlated effects are present, incorporating them yields substantial gains. On average, the Mundlak specification reduces MSFE by about 3.9% relative to the baseline RE predictor (Chamberlain: 1.3%). Using the *wrong* CE specification leads only to minor losses. Overall, these findings suggest that incorporating CE is beneficial in practice, with the Mundlak specification providing a particularly robust choice.

4.4.4 Summary of Simulation Results

The following general results emerge from the simulation study:

- Applying forecasting methods that assume stationary initial conditions to non-stationary data leads to substantial losses in forecast accuracy, whereas the reverse misspecification is less severe. In the former case, forecasts rely on systematically biased moment information, resulting in pronounced errors, particularly when the time dimension is short.

- When additional covariates are included, the only correctly specified model is based on non-stationary starting conditions. Models derived under stationary IC do not fully remove lagged components, leading to misspecification that can offset the gains from conditioning on additional regressors.

- Shrinkage gains are largest when heterogeneity is moderate. The larger the cross-sectional dispersion, the more of the true signal is shrunk toward the mean by RE

models, reducing their relative advantage over FE alternatives.

- Incorporating a correlated-effects framework is essentially risk-free and can substantially improve forecast accuracy when individual effects are correlated with the regressors. The Mundlak specification in particular provides a robust choice. The Chamberlain specification can be advantageous when the informational content of regressors varies over time, e.g., when recent observations are more informative. Regularization can help stabilize the estimation of the Chamberlain projection.

- All methods deteriorate as persistence approaches unity. Since many second-moment expressions scale with $(1-\rho)^{-1}$, near–unit-root behavior amplifies estimation errors. In highly persistent environments, differencing can yield a more stable estimation problem, offsetting the loss of the first observation.

- Forecast performance is strongly related to the length of the time series. As T increases, differences across forecasting methods narrow substantially, and for sufficiently long panels most approaches approximate similar forecasting rules.

- The role of the cross-sectional dimension is particularly important when additional covariates are included, as it determines the precision with which ρ and β are estimated. However, most gains from pooling are realized once N reaches moderate values, with limited additional improvements beyond $N \geqslant 100$.

- Across all data-generating processes considered, the RE predictor under non-stationary initial conditions emerges as the most reliable method, combining strong forecast accuracy with robustness to misspecification and estimation error.

4.5 Empirical Application

4.5.1 Data

The empirical application uses German firm-level productivity data. Productivity is measured as revenue per employee. The analysis is based on a balanced panel of $N = 10{,}580$ firms observed over $T = 13$ years (2012 to 2024). The outcome variable is defined as firm-level productivity expressed as a deviation from the industry-level mean. Average firm size and the federal state of operation are considered as potential explanatory variables.

An important practical feature of the empirical application is that the available explanatory variables are time-invariant at the firm level. Firm size is measured by its time average,

reflecting the fact that within-firm variation is negligible over the sample period. The federal state indicator is constant because firms did not relocate.[5] As a consequence, the regressors enter the model exclusively through their long-run component like in the Mundlak-CE specification:

$$\lambda_i = \bar{x}_i'\gamma + \eta_i.$$

Figure C.1 in Appendix C.4 shows the distribution of firm-level productivity and its relationship with observable firm characteristics. The distribution of mean productivity across firms exhibits substantial dispersion. A simple variance decomposition shows that roughly 83% of total variation is attributable to persistent differences in firm-specific mean productivity, while within-firm fluctuations account for a comparatively small share of approximately 17%. This implies that productivity differences across firms are highly persistent and dominate short-run dynamics. As a result, forecasting performance is likely to depend on how individual heterogeneity is modeled.

To assess persistence, a dynamic panel model with firm fixed-effects is estimated. A bias-corrected estimator yields $\hat{\rho} = 0.5335$, indicating persistent dynamics that lie within the ρ-grid of the simulation analysis. To determine whether the available covariates help explain persistent productivity differences, we regress firm-level mean productivity on average firm size and federal-state indicators. Both covariates are statistically significant and jointly explain a statistically significant share of the cross-sectional variation in mean productivity. While the overall fit is modest ($R^2 = 0.0274$), this implies that observable characteristics capture only part of the persistent heterogeneity, leaving substantial scope for modeling unobserved individual effects. In addition, the boxplots reveal well-known regional patterns: firms located in eastern German states tend to exhibit lower productivity, while firms in city states are more productive on average (compare, e.g., Roehl, 2025).

4.5.2 Results

To reduce the influence of *one-panel luck* and to mirror the simulation study, a design-based Monte Carlo procedure is employed in which the forecast target is fixed while the length of the training sample varies. For training sample lengths of $T \in \{6, 8, 10, 12\}$, $N = 200$ firms are randomly drawn from the population without replacement. Competing forecasting models are estimated on the resulting subsamples, and one-step-ahead MSFE for the productivity deviation from the industry mean in 2024 is computed.[6] This procedure is repeated 1,000

[5] More specifically, if a firm relocates, it enters the panel as a new panel unit. Therefore, at the panel-unit-level, relocations are excluded by construction.

[6] The forecasting objective is always the prediction of productivity deviation from the industry mean in 2024. For $T < 12$, only the T most recent observations prior to 2024 are used for training.

times for each training sample length.

Tables 4.4 and 4.5 report the estimated common parameters and the corresponding MSFE for the competing forecasting models. Density estimates of the coefficients are provided in Appendix C.4 (Figure C.2).[7]

Table 4.4: Average Estimates of Common Parameters across Training Sample Lengths

Time	$\hat{\rho}$	$\hat{\sigma}^2_\varepsilon$	$\hat{\sigma}^2_\lambda$	$\hat{\sigma}^2_\eta$	$\hat{\sigma}^2_{\mu(\lambda)}$	$\hat{\sigma}^2_{\mu(\eta)}$
$T = 6$	0.2759	0.0825	0.2914	0.2611	0.5008	0.4478
$T = 8$	0.3045	0.0763	0.2602	0.2325	0.5011	0.447
$T = 10$	0.325	0.0746	0.2391	0.2142	0.4923	0.4405
$T = 12$	0.3533	0.0801	0.2157	0.1923	0.4929	0.4393

The table shows that estimated persistence is moderate and tends to increase with the length of the training sample. The estimated innovation variance $\hat{\sigma}^2_\varepsilon$ is virtually identical across training sample lengths and invariant to the inclusion of time-invariant covariates. This reflects the fact that $\hat{\sigma}^2_\varepsilon$ is identified from the within-firm variation of the residual process $z_{it} = y_{it} - \hat{\rho} y_{i,t-1}$. Since time-invariant covariates affect only the cross-sectional mean of z_{it}, they do not enter the within transformation underlying the Swamy–Arora estimator. By contrast, including time-constant covariates affects the decomposition of between-firm variation: observable firm characteristics explain part of the persistent heterogeneity, reducing the variance of the remaining unobserved component σ^2_η relative to σ^2_λ.

Table 4.5 reports the MSFE across methods and training sample lengths. All models are estimated with and without covariates. Note that for the simple FE predictors, including time-invariant covariates cannot improve forecasts as both rules depend only on unit-specific sample means, so that $\bar{x}_i'\gamma$ cancels out. The same logic applies to the FE_D estimator, which is based on first differences. We observe that models with explanatory variables achieve uniformly lower MSFE. Further, most models outperform the simple mean benchmark, which forecasts $y_{i,T+1} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$.

[7] Densities are shown for the longest training sample ($T = 12$). The qualitative patterns are identical for shorter training periods.

Table 4.5: MSFE across forecasting methods and training sample lengths. The MSFE of the best-performing model per experiment is shown in bold. For further notes (abbreviations), see Table 4.1.

Time	Covariates	Mean	FE(S)	FE(N)	$\mathbf{FE_D(S)}$	RE(S)	RE(N)	$\mathbf{TW_P(N)}$
$T = 6$	Yes	0.1262	0.1124	0.1108	0.1123	0.1111	**0.1089**	0.1302
	No					0.1113	0.1091	0.1328
$T = 8$	Yes	0.1343	0.1095	0.1057	0.109	0.1081	**0.1046**	0.1162
	No					0.1083	0.1048	0.1168
$T = 10$	Yes	0.1484	0.1136	0.1082	0.1131	0.1122	**0.1074**	0.1131
	No					0.1124	0.1075	0.1137
$T = 12$	Yes	0.164	0.1193	0.1141	0.1187	0.1177	**0.1132**	0.1179
	No					0.1179	0.1134	0.1181

Overall, the empirical findings closely align with the patterns documented in the simulation study. The $RE(N)$ predictor consistently delivers the strongest performance, achieving a 23.6% improvement relative to the mean benchmark (averaged over all training sample lengths). Methods derived under stationary initial conditions perform worse (RE : 21.0%, FE : 20.0%), while the Tweedie estimator exhibits comparatively weaker accuracy (TW_p : 15.2%). These results reinforce the importance of modeling individual heterogeneity through shrinkage and accounting properly for initial conditions.

To assess the practical usefulness of the proposed decay test, we implemented a test-based model selection rule that chooses $RE(N)$ when the null of stationary initial conditions is rejected and $RE(S)$ otherwise. The null was rejected in 95% of the replications, indicating strong evidence in favor of non-stationary initialization in this environment. Consequently, the performance of the test-based rule (23.4% improvement relative to the mean benchmark) is virtually identical to that of the unconditional $RE(N)$ predictor. While the test does not generate additional gains in this application, it confirms the relevance of the non-stationary specification without deteriorating forecast performance.

4.6 Conclusion

Panel data econometrics has traditionally focused on the consistent estimation and inference of marginal effects, while panel forecasting has received comparatively little attention. Although standard time-series methods exploit temporal dependence effectively, they often make limited use of the cross-sectional information and forecasting is still largely treated as a time-series problem.

Assuming that common parameters can be estimated consistently using both the cross-sectional and time-series dimensions, this chapter focuses on the role of individual hetero-

geneity in panel forecasting. It develops shrinkage-based forecasting methods to improve the external validity of the forecast and compares their performance to a recent empirical Bayes approach for panel forecasting (Tweedie). Particular attention is given to the treatment of initial conditions. Conditioning on the full time series is appropriate under stationarity, while quasi-differencing provides a more robust alternative when the process has not yet converged. The simulation results demonstrate that misspecifying the initial conditions can lead to substantial losses in forecast accuracy, whereas methods derived under non-stationary initial conditions remain comparatively robust. Across all data-generating processes considered, the random-effects predictor based on non-stationary initial conditions (and possibly extended by Mundlak-CE) emerges as the most reliable approach. More generally, forecast performance depends critically on the degree of heterogeneity, persistence, and the sample dimensions: shrinkage gains are strongest in short panels with moderate heterogeneity, but diminish as cross-sectional dispersion or the time dimension increases. Near–unit-root behavior amplifies estimation error and limits the benefits from pooling beyond moderate cross-sectional sizes.

Several directions for future research follow naturally from this analysis. One extension would be to investigate whether shrinkage techniques developed for individual intercepts can also be applied to common coefficients. Further progress could be made by allowing for heterogeneous initial conditions across units and by developing more elaborate testing procedures to detect such patterns. Moreover, the presented approaches rely on the availability of a sufficiently large cross section, which could be exploited more systematically. Strategies based on sample splitting or local learning, similar in spirit to matching procedures, may help improve panel forecasting methods further.

A Appendix for Chapter 2

A.1 The Limit of REGSC

For $\lambda_1 \to \infty$ and $\lambda_2 \to \infty$ the objective function reduces to

$$Q(\lambda_1, \lambda_2) = \lambda_1 w'w + \lambda_2(1 - \mathbf{1}'w)^2$$

The derivative is obtained as

$$\frac{\partial Q(\lambda_1, \lambda_2)}{\partial w} = 2\lambda_1 w - 2\lambda_2(\mathbf{1} - \mathbf{1}\mathbf{1}'w)$$

By setting the derivative to zero and multiplying with $\mathbf{1}$ we obtain:

$$\lambda_1 \mathbf{1}'w - \lambda_2(J - J\mathbf{1}'w) = 0$$

where $\mathbf{1}'w = \sum w_i$. Solving for $\mathbf{1}'w$ we obtain

$$\mathbf{1}'w = \frac{J}{J + \lambda_1/\lambda_2}$$

and due to the symmetry of the objective function with respect to the elements of the weight vector we have

$$w_i = \frac{1}{J + \frac{\lambda_1}{\lambda_2}}$$

A.2 Derivation of the REGSC Prior

The objective function of the REGSC estimator is given by

$$\begin{aligned} Q(w,\lambda_1,\lambda_2) &= \sum_{t=1}^{T_0} \underbrace{\left(y_{0,t} - \mu^* - \sum_{j=1}^{J} w_j y_{j,t}\right)^2}_{RSS} + \lambda_1 \underbrace{\left(\sum_{j=1}^{J} w_j^2\right)}_{Ridge} + \lambda_2 \underbrace{\left(1 - \sum_{j=1}^{J} w_j\right)^2}_{weight\ penalty} + C \\ &= RSS + \lambda_1 w'w + \lambda_2 (w - \iota_J)' \mathbf{1_J 1_J'} (w - \iota_J) + C \end{aligned}$$

where ι_J is a $J \times 1$ vector with identical elements $1/J$. The additional constant C is just introduced for convenience and does not affect the solution. We are looking for a Bayesian interpretation of the likelihood of the form

$$l(w,\lambda_1,\lambda_2) = const - \frac{1}{2\sigma^2} RSS - \frac{1}{2}(w-\mu_0)' V_0^{-1} (w-\mu_0)$$

where the second term represents the prior distribution. From the squared w_i, we find the covariance matrix:

$$\begin{aligned} \frac{1}{\sigma^2} V_0^{-1} &= \lambda_1 I_J + \lambda_2 \mathbf{1_J 1_J'} \\ &= \lambda_1 (I_J - P) + (\lambda_1 + J\lambda_2) P \quad \text{with } P = \frac{1}{J} \mathbf{1_J 1_J'}, \end{aligned}$$

and inverting this relation yields

$$\begin{aligned} \frac{1}{\sigma^2} V_0 &= \frac{1}{\lambda_1}(I_J - P) + \frac{1}{\lambda_1 + J\lambda_2} P \\ &= \frac{1}{\lambda_1} I_J - \frac{\lambda_2}{\lambda_1(\lambda_1 + J\lambda_2)} \mathbf{1_J 1_J'}. \end{aligned}$$

In a similar manner we obtain

$$-\frac{2}{\sigma^2} w' V_0^{-1} \mu_0 = -2\lambda_2 w' \mathbf{1_J 1_J'} \iota_J$$

and, therefore,

$$\begin{aligned} \frac{1}{\sigma^2} V_0^{-1} \mu_0 &= \lambda_2 \mathbf{1_J 1_J'} \iota_J \\ \mu_0 &= \lambda_2 \sigma^2 V_0 \mathbf{1_J 1_J'} \iota_J \\ &= \lambda_2 / (\lambda_1 + J\lambda_2). \end{aligned}$$

This gives the parameter for the prior distribution. Note that the constant C ensures that also the constant term will match.

A.3 Derivation of the Analytical REGSC Distribution

It has already been shown that the REGSC estimator has the following closed form solution:

$$\widehat{w}_{\lambda_1,\lambda_2} = \left(\widetilde{Z}'\widetilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J'\right)^{-1} \left(\widetilde{Z}'\widetilde{y_0} + \lambda_2 \mathbf{1}_J\right).$$

Similar to the variance of the OLS estimator, the variance of the REGSC estimator conditional on the explanatory variables is derived:

$$\begin{aligned}
V(\widehat{w}_{\lambda_1,\lambda_2}|\widetilde{Z}) &= (\widetilde{Z}'\widetilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1}\widetilde{Z}'\widetilde{Z} \;\; V(\widehat{w}_{\lambda_1,\lambda_2}|\widetilde{Z}) \;\; \left[(\widetilde{Z}'\widetilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1}\widetilde{Z}'\widetilde{Z}\right]' \\
&= (\widetilde{Z}'\widetilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1}\widetilde{Z}'\widetilde{Z} \;\; V(\widehat{w}_{\lambda_1,\lambda_2}|\widetilde{Z}) \;\; \widetilde{Z}'\widetilde{Z}(\widetilde{Z}'\widetilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1} \\
&= (\widetilde{Z}'\widetilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1}\widetilde{Z}'\widetilde{Z} \;\; \sigma^2(\widetilde{Z}'\widetilde{Z})^{-1}\widetilde{Z}'\widetilde{Z}(\widetilde{Z}'\widetilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1} \\
&= \sigma^2(\widetilde{Z}'\widetilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1}\widetilde{Z}'\widetilde{Z}(\widetilde{Z}'\widetilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1}
\end{aligned}$$

For a normally distributed error term, it follows that the REGSC estimator is normally distributed with $\widehat{w}_{\lambda_1,\lambda_2} \sim \mathcal{N}\left(\widehat{w}_{\lambda_1,\lambda_2}, V(\widehat{w}_{\lambda_1,\lambda_2})\right)$.

A.4 VAR(2) Representation of Differenced VAR(1) Process

Consider a VAR model of order 1 (of dimension k):

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{u}_t,$$

where:

- $\mathbf{A}_1$ is the $k \times k$ coefficient matrix,
- $\mathbf{u}_t$ is a white noise vector with mean zero and covariance matrix Σ_u.

Define the cumulative series $\mathbf{z}_t$ as:

$$\mathbf{z}_t = \sum_{i=1}^{t} \mathbf{y}_i.$$

This implies:

$$\mathbf{z}_t = \mathbf{z}_{t-1} + \mathbf{y}_t.$$

Substituting the VAR(1) model for $\mathbf{y}_t$, we obtain:

$$\mathbf{z}_t = \mathbf{z}_{t-1} + \mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{u}_t.$$

Using the relationship $\mathbf{y}_{t-1} = \mathbf{z}_{t-1} - \mathbf{z}_{t-2}$, we rewrite:

$$\mathbf{z}_t = \mathbf{z}_{t-1} + \mathbf{A}_1(\mathbf{z}_{t-1} - \mathbf{z}_{t-2}) + \mathbf{u}_t.$$

Simplifying gives:

$$\mathbf{z}_t = (\mathbf{I} + \mathbf{A}_1)\mathbf{z}_{t-1} - \mathbf{A}_1 \mathbf{z}_{t-2} + \mathbf{u}_t.$$

The above equation is a VAR(2) model for the cumulative series $\mathbf{z}_t$:

$$\mathbf{z}_t = \mathbf{B}_1 \mathbf{z}_{t-1} + \mathbf{B}_2 \mathbf{z}_{t-2} + \mathbf{u}_t,$$

where:

$$\mathbf{B}_1 = \mathbf{I} + \mathbf{A}_1,$$
$$\mathbf{B}_2 = -\mathbf{A}_1.$$

The cumulative nature of $\mathbf{z}_t$ introduces implicit time trends, as cumulative sums often exhibit

growth or trend-like behavior over time. These trends align naturally with the VAR(2) framework, where they emerge from the dynamics of the lagged cumulative terms.

A.5 Static Data Simulation Results

Table A.1: Average RMSE in SC-Style DGP

T_{pre}	**Method**	**5**	**10**	**15**	**20**	**25**	**30**
$T_{pre} = 20$	SC	1.0157	1.0544	1.0692	1.0600	1.0684	1.0770
	AUGSC_{Ridge}	1.0447	1.0668	1.1089	2.3767	2.1845	1.7840
	OLS	1.1510	1.4755	2.2826	5.8202	–	–
	REGSC	1.1025	1.1059	1.0956	1.0869	1.0754	1.0712
	NET	1.1048	1.1474	1.1446	1.1286	1.1360	1.1334
	FACTOR	1.1344	1.1004	1.0980	1.0712	1.0782	1.0693
$T_{pre} = 50$	SC	0.9946	1.0165	1.0198	1.0353	1.0285	1.0315
	AUGSC_{Ridge}	0.9989	1.0255	1.0292	1.0479	1.0428	1.0442
	OLS	1.0388	1.1117	1.1854	1.2905	1.4184	1.5971
	REGSC	1.0394	1.0481	1.0343	1.0458	1.0257	1.0340
	NET	1.0311	1.0545	1.0562	1.0616	1.0519	1.0576
	FACTOR	1.1006	1.0483	1.0410	1.0432	1.0335	1.0381
$T_{pre} = 100$	SC	0.9720	1.0008	1.0002	1.0081	0.9973	1.0211
	AUGSC_{Ridge}	0.9730	1.0047	1.0056	1.0147	1.0050	1.0281
	OLS	0.9908	1.0339	1.0581	1.0966	1.1301	1.1791
	REGSC	0.9907	1.0195	1.0068	1.0156	0.9977	1.0192
	NET	0.9871	1.0182	1.0135	1.0254	1.0096	1.0290
	FACTOR	1.0742	1.0387	1.0100	1.0230	1.0025	1.0216

$1,000$ Iterations per Donor/Method combination, $T_{pre} \in \{20, 50, 100\}$ and $T_{post} = 10$.

Table A.2: Average RMSE in Factor-Style DGP

T_{pre}	**Method**	**5**	**10**	**15**	**20**	**25**	**30**
$T_{pre} = 20$	SC	1.4198	1.2700	1.2321	1.2145	1.1883	1.1632
	AUGSC_{Ridge}	1.4115	1.2718	1.2125	1.5652	1.2030	1.1731
	OLS	1.3523	1.5895	2.4407	6.9961	–	–
	REGSC	1.3013	1.2059	1.2020	1.1900	1.1841	1.1811
	NET	1.2843	1.2397	1.2422	1.2278	1.2094	1.1989
	FACTOR	1.2399	1.1372	1.1081	1.0950	1.0986	1.0936
$T_{pre} = 50$	SC	1.3912	1.2431	1.1862	1.1386	1.1481	1.1368
	AUGSC_{Ridge}	1.3871	1.2036	1.1533	1.1307	1.1159	1.1171
	OLS	1.1878	1.2130	1.2498	1.3253	1.4637	1.6523
	REGSC	1.1823	1.1355	1.1015	1.0770	1.0803	1.0917
	NET	1.1721	1.1501	1.1136	1.0915	1.0985	1.1120
	FACTOR	1.1521	1.1000	1.0637	1.0425	1.0369	1.0506
$T_{pre} = 100$	SC	1.3860	1.2054	1.1566	1.1220	1.1247	1.1025
	AUGSC_{Ridge}	1.3627	1.1766	1.1348	1.0964	1.1013	1.0828
	OLS	1.1616	1.1185	1.1392	1.1365	1.1845	1.1948
	REGSC	1.1621	1.0905	1.0843	1.0461	1.0606	1.0381
	NET	1.1571	1.0963	1.0884	1.0595	1.0766	1.0537
	FACTOR	1.1447	1.0673	1.0589	1.0223	1.0390	1.0199

$1,000$ Iterations per Donor/Method combination, $T_{pre} \in \{20, 50, 100\}$ and $T_{post} = 10$.

A.6 Dynamic Data Simulation Results

Table A.3: Average RMSE in Dynamic DGPs for $T_{pre} = 20$

DGP	Method	5	10	15	20
Case 3:	SC	2.6800	2.5240	2.4092	2.3818
Stationary	AUGSC_{Ridge}	2.7091	2.5330	2.4688	3.3823
	OLS	2.7845	3.0724	4.4832	8.9203
	REGSC	2.6719	2.5503	2.4672	2.4249
	NET	2.6584	2.5097	2.4155	2.3810
	FACTOR	2.7421	2.5886	2.5235	2.4971
	VAR	–	–	–	–
	REGVAR	2.7598	2.6638	2.5189	2.4570
	NETVAR	2.7088	2.4865	2.4275	2.3928
Case 4:	SC	2.8816	3.2696	3.6575	5.0973
Non-Stationary	AUGSC_{Ridge}	3.0312	3.3823	3.7974	7.6195
	OLS	3.3242	4.3165	6.2506	17.9334
	REGSC	2.9984	3.3647	3.7672	5.1819
	NET	3.0802	3.4592	3.7035	4.9259
	FACTOR	3.0937	3.5576	3.8997	5.4014
	VAR	–	–	–	–
	REGVAR	3.2150	3.3911	3.6663	4.9437
	NETVAR	2.9918	3.3044	3.5859	4.7940
Case 5:	SC	3.0466	2.7271	2.7706	2.7875
Cointegrated	AUGSC_{Ridge}	3.3287	2.8044	2.8618	3.6898
	OLS	3.4605	3.4409	4.6626	9.9770
	REGSC	3.0514	2.719	2.7028	2.7441
	NET	3.2416	2.8164	2.7638	2.7448
	FACTOR	3.1902	2.8348	2.8875	2.9052
	VAR	–	–	–	–
	REGVAR	3.1683	2.7062	2.7389	2.7411
	NETVAR	2.9793	2.7169	2.7001	2.7075

$1,000$ Iterations per Donor/Method combination, $T_{pre} = 20$, $T_{post} = 10$

Table A.4: Average RMSE in Dynamic DGPs for $T_{pre} = 50$

DGP	**Method**	**5**	**10**	**15**	**20**
Case 3:	SC	2.5204	2.4077	2.3037	2.2944
Stationary	AUGSC_{Ridge}	2.5203	2.3739	2.2835	2.2286
	OLS	2.4014	2.3429	2.2646	2.2887
	REGSC	2.4216	2.3222	2.2014	2.1806
	NET	2.3871	2.2662	2.1681	2.1149
	FACTOR	2.5543	2.4489	2.3920	2.3799
	VAR	2.4199	2.4434	–	–
	REGVAR	2.3784	2.2547	2.1534	2.0809
	NETVAR	2.3267	2.1895	2.0298	1.9402
Case 4:	SC	2.8039	3.1841	3.8007	4.9259
Non-Stationary	AUGSC_{Ridge}	2.8876	3.2442	3.7707	4.6865
	OLS	3.0289	3.4386	4.0072	4.7709
	REGSC	2.8521	3.2295	3.7673	4.6051
	NET	2.9461	3.2792	3.7462	4.4666
	FACTOR	2.9905	3.4189	4.0811	5.2793
	VAR	2.8412	4.0032	–	–
	REGVAR	2.7067	2.9752	3.4548	4.0217
	NETVAR	2.5729	2.7447	3.0266	3.4661
Case 5:	SC	2.7544	2.8205	2.6919	2.6239
Cointegrated	AUGSC_{Ridge}	2.9998	2.9281	2.6164	2.4738
	OLS	3.0104	3.0065	2.7025	2.6271
	REGSC	2.8172	2.7848	2.5793	2.4303
	NET	2.9366	2.8513	2.5431	2.4256
	FACTOR	2.8767	2.9378	2.7902	2.7201
	VAR	2.8618	4.0157	–	–
	REGVAR	2.6437	2.6657	2.5591	2.3888
	NETVAR	2.5540	2.4996	2.3415	2.2779

$1,000$ Iterations per Donor/Method combination, $T_{pre} = 50$, $T_{post} = 10$

Table A.5: Average RMSE in Dynamic DGPs for $T_{pre} = 100$

DGP	**Method**	**5**	**10**	**15**	**20**
Case 3:	SC	2.4980	2.3463	2.2805	2.2014
Stationary	AUGSC_{Ridge}	2.4863	2.3141	2.1941	2.0821
	OLS	2.3072	2.1604	2.0448	1.9487
	REGSC	2.3186	2.1786	2.0495	1.9615
	NET	2.3065	2.1508	2.0230	1.9224
	FACTOR	2.5233	2.4029	2.3380	2.3137
	VAR	2.2621	2.0302	1.7784	–
	REGVAR	2.2543	2.0429	1.7713	1.5501
	NETVAR	2.2292	1.9926	1.7406	1.4465
Case 4:	SC	2.7061	3.2894	3.7263	4.8428
Non-Stationary	AUGSC_{Ridge}	2.8124	3.3643	3.6788	4.4557
	OLS	2.9424	3.4631	3.7515	4.3765
	REGSC	2.8273	3.3116	3.6526	4.3478
	NET	2.8982	3.3791	3.6401	4.2997
	FACTOR	2.8669	3.4590	3.9994	5.1887
	VAR	2.3290	2.4378	2.3673	–
	REGVAR	2.4021	2.5754	2.5657	2.5931
	NETVAR	2.2914	2.3680	2.2490	2.0841
Case 5:	SC	2.7439	2.8006	2.6779	2.6771
Cointegrated	AUGSC_{Ridge}	3.1587	3.0129	2.6715	2.4754
	OLS	3.0471	3.0516	2.6955	2.4465
	REGSC	2.9070	2.8560	2.6190	2.4289
	NET	2.9911	2.9547	2.5989	2.4041
	FACTOR	2.8165	2.8976	2.8079	2.7231
	VAR	2.3556	2.4160	2.2410	–
	REGVAR	2.4115	2.3714	2.2037	2.0396
	NETVAR	2.3227	2.2564	1.9798	1.8555

$1,000$ Iterations per Donor/Method combination, $T_{pre} = 100$, $T_{post} = 10$

Table A.6: Median RMSE in Dynamic DGPs for $T_{pre} = 20$

DGP	Method	5	10	15	20
Case 3:	SC	2.3146	2.2551	1.9923	1.9479
Stationary	AUGSC_{Ridge}	2.3339	2.2923	2.0597	2.5407
	OLS	2.4173	2.7623	3.9500	7.2937
	REGSC	2.3194	2.3681	2.1622	2.0999
	NET	2.2668	2.1713	2.0333	1.9763
	FACTOR	2.4055	2.4163	2.1595	2.1941
	VAR	–	–	–	–
	REGVAR	2.4204	2.4926	2.1980	2.1258
	NETVAR	2.4138	2.2299	2.0852	2.0089
Case 4:	SC	2.6919	3.0263	3.4057	4.5223
Non-Stationary	AUGSC_{Ridge}	2.7938	3.0287	3.4654	6.2593
	OLS	3.0049	3.7409	5.5244	15.0597
	REGSC	2.7184	3.1473	3.3564	4.7937
	NET	2.8034	3.0834	3.2476	4.3633
	FACTOR	2.8849	3.2791	3.5898	5.0209
	VAR	–	–	–	–
	REGVAR	2.8739	3.1306	3.4137	4.4186
	NETVAR	2.7834	2.9674	3.3118	4.4241
Case 5:	SC	2.6820	2.4398	2.5491	2.4532
Cointegrated	AUGSC_{Ridge}	2.8817	2.5139	2.5766	3.0951
	OLS	3.0782	2.9764	4.0203	7.8101
	REGSC	2.6716	2.3731	2.3944	2.4583
	NET	2.8537	2.4199	2.3402	2.4248
	FACTOR	2.8602	2.5143	2.5514	2.6350
	VAR	–	–	–	–
	REGVAR	2.7435	2.4186	2.3980	2.4484
	NETVAR	2.6170	2.4161	2.3646	2.3181

$1,000$ Iterations per Donor/Method combination, $T_{pre} = 20$, $T_{post} = 10$

Table A.7: Median RMSE in Dynamic DGPs for $T_{pre} = 50$

DGP	Method	5	10	15	20
Case 3:	SC	2.1527	2.1128	1.9197	1.9710
Stationary	AUGSC_{Ridge}	2.1497	2.1035	1.9011	1.9146
	OLS	2.1800	2.0288	1.8947	2.0360
	REGSC	2.1923	2.0653	1.9646	1.9565
	NET	2.1994	1.9888	1.8108	1.8741
	FACTOR	2.3318	2.2050	2.1286	2.1874
	VAR	2.1497	2.1288	–	–
	REGVAR	2.1634	1.9821	1.8235	1.8422
	NETVAR	2.0857	1.9188	1.7203	1.7018
Case 4:	SC	2.5190	2.8811	3.4963	4.4441
Non-Stationary	AUGSC_{Ridge}	2.6085	2.9355	3.4057	4.2223
	OLS	2.6671	2.9841	3.6348	4.2033
	REGSC	2.5390	2.8985	3.3985	4.0976
	NET	2.5918	2.8696	3.3237	3.9499
	FACTOR	2.6975	3.1046	3.7211	4.8242
	VAR	2.4447	3.3942	–	–
	REGVAR	2.4081	2.6670	3.1888	3.5930
	NETVAR	2.2490	2.4301	2.6256	3.0131
Case 5:	SC	2.5103	2.5179	2.3388	2.3225
Cointegrated	AUGSC_{Ridge}	2.6708	2.6073	2.2763	2.2002
	OLS	2.6908	2.6687	2.3472	2.2895
	REGSC	2.5005	2.4945	2.2357	2.1547
	NET	2.6324	2.5278	2.2266	2.1537
	FACTOR	2.5910	2.6289	2.4927	2.4405
	VAR	2.5016	3.3950	–	–
	REGVAR	2.3783	2.4112	2.2524	2.1621
	NETVAR	2.2902	2.1329	2.0571	2.0073

$1,000$ Iterations per Donor/Method combination, $T_{pre} = 50$, $T_{post} = 10$

Table A.8: Median RMSE in Dynamic DGPs for $T_{pre} = 100$

DGP	Method	5	10	15	20
Case 3:	SC	2.2279	2.0877	1.9111	2.0078
Stationary	AUGSC_{Ridge}	2.2235	2.0632	1.8427	1.8655
	OLS	2.0815	1.9519	1.8073	1.7365
	REGSC	2.0977	1.9734	1.8084	1.7643
	NET	2.0775	1.9333	1.7753	1.7203
	FACTOR	2.3177	2.2151	2.1029	2.0791
	VAR	2.0221	1.8079	1.4739	–
	REGVAR	1.9862	1.8004	1.4974	1.2707
	NETVAR	1.9968	1.7793	1.4671	1.2348
Case 4:	SC	2.4528	2.9918	3.3401	4.4370
Non-Stationary	AUGSC_{Ridge}	2.5347	2.9693	3.3456	3.9513
	OLS	2.6420	3.0877	3.3858	3.8247
	REGSC	2.5616	3.0066	3.2685	3.8499
	NET	2.5915	3.0350	3.2863	3.8033
	FACTOR	2.6450	3.1572	3.5251	4.7315
	VAR	2.0213	2.0957	2.0785	–
	REGVAR	2.1180	2.2641	2.2471	2.2469
	NETVAR	1.9863	2.0747	1.9639	1.7419
Case 5:	SC	2.5629	2.4752	2.3726	2.3840
Cointegrated	AUGSC_{Ridge}	2.8415	2.6694	2.2539	2.1521
	OLS	2.6722	2.5965	2.3362	2.1750
	REGSC	2.5618	2.4627	2.3019	2.1576
	NET	2.6176	2.5413	2.2502	2.1194
	FACTOR	2.5631	2.6457	2.5031	2.3740
	VAR	2.0599	2.0380	1.9858	–
	REGVAR	2.1882	2.0047	1.9874	1.7131
	NETVAR	2.0371	1.9677	1.7428	1.5675

$1,000$ Iterations per Donor/Method combination, $T_{pre} = 100$, $T_{post} = 10$

A.7 Application

Table A.9: Estimation Results: Tobacco Control Program

Type	**Method**	$\boldsymbol{R^2}$	$\boldsymbol{\delta_{end}}$	$\bar{\boldsymbol{\delta}}$	$\sum \boldsymbol{\delta}$	p**-value**
Levels	SC	0.9764	-25.58	-18.84	-226.09	0.026
	REGSC	1.0000	-21.51	-14.52	-174.23	0.026
	REGVAR	0.9983	-23.68	-14.51	-174.08	0.179
	NET	0.9984	-24.21	-16.36	-196.32	0.026
	NETVAR	0.9993	-22.13	-16.63	-199.50	0.103
	FACTOR	0.9959	-28.04	-18.73	-224.76	0.051
1st Diff.	SC	0.8338	-38.05	-29.03	-348.39	0.359
	REGSC	0.9914	-14.43	-11.45	-137.41	0.205
	REGVAR	0.9995	-15.42	-10.01	-120.12	0.103
	NET	0.9938	-13.49	-11.44	-137.27	0.205
	NETVAR	0.9999	-12.55	-8.64	-103.70	0.128
	FACTOR	0.9943	-12.28	-11.36	-136.35	0.128

B Appendix for Chapter 3

B.1 DMA in Comparison to Alternative Methods

Table B.1: Static Case: The table shows the RMSE/ MAE that result from the three time series lengths $T = \{60, 120, 360\}$ and the four distinguished sub-cases (2 resp. 4 explanatory variables/ 0 resp. 2 noise variables). Bold errors indicate the best performing model per row.

error	obs	n_{stat}	n_{noise}	O_{common}	$O_{r-specific}$	NO_{common}	$NO_{r-specific}$	recGAM	recOLS	rolOLS	naïve
RMSE	60	2	0	0.388	0.390	0.392	0.403	0.523	**0.386**	0.394	0.708
RMSE	60	2	2	**0.382**	0.385	0.395	0.402	0.794	0.398	0.426	0.706
RMSE	60	4	0	0.413	**0.413**	0.428	0.438	0.792	0.427	0.436	0.830
RMSE	60	4	2	**0.414**	0.414	0.432	0.441	1.174	0.442	0.479	0.834
RMSE	120	2	0	**0.373**	0.373	0.375	0.378	0.413	0.376	0.373	0.708
RMSE	120	2	2	**0.370**	0.371	0.375	0.376	0.489	0.381	0.386	0.703
RMSE	120	4	0	0.391	**0.388**	0.396	0.396	0.484	0.409	0.393	0.837
RMSE	120	4	2	0.391	**0.389**	0.399	0.398	0.599	0.416	0.408	0.841
RMSE	360	2	0	0.362	**0.361**	0.363	0.362	0.370	0.371	0.363	0.709
RMSE	360	2	2	0.361	**0.360**	0.364	0.362	0.386	0.373	0.366	0.709
RMSE	360	4	0	0.371	**0.368**	0.373	0.370	0.389	0.402	0.375	0.845
RMSE	360	4	2	0.371	**0.368**	0.375	0.371	0.407	0.404	0.379	0.844
MAE	60	2	0	0.310	0.312	0.314	0.322	0.385	**0.309**	0.315	0.571
MAE	60	2	2	**0.306**	0.308	0.316	0.322	0.536	0.319	0.341	0.567
MAE	60	4	0	0.330	**0.330**	0.342	0.349	0.538	0.341	0.348	0.668
MAE	60	4	2	**0.331**	0.331	0.346	0.352	0.767	0.353	0.380	0.673
MAE	120	2	0	**0.297**	0.298	0.299	0.301	0.323	0.300	0.298	0.568
MAE	120	2	2	0.296	**0.296**	0.299	0.300	0.373	0.304	0.308	0.563
MAE	120	4	0	0.312	**0.310**	0.316	0.316	0.372	0.327	0.314	0.670
MAE	120	4	2	0.312	**0.311**	0.318	0.318	0.440	0.332	0.325	0.673
MAE	360	2	0	0.289	**0.288**	0.289	0.289	0.295	0.296	0.290	0.566
MAE	360	2	2	0.288	**0.287**	0.290	0.289	0.307	0.298	0.292	0.567
MAE	360	4	0	0.296	**0.294**	0.298	0.295	0.309	0.320	0.299	0.675
MAE	360	4	2	0.296	**0.294**	0.299	0.296	0.323	0.322	0.302	0.674

Table B.2: Dynamic Case: For further notes see Table B.1.

error	obs	n_{stat}	n_{noise}	O_{common}	$O_{r-specific}$	NO_{common}	$NO_{r-specific}$	recGAM	recOLS	rolOLS	naïve
RMSE	60	2	0	**0.388**	0.392	0.394	0.407	0.509	0.388	0.391	0.700
RMSE	60	2	2	**0.380**	0.383	0.394	0.402	0.781	0.398	0.420	0.696
RMSE	60	4	0	**0.389**	0.393	0.411	0.421	0.771	0.420	0.429	0.794
RMSE	60	4	2	**0.388**	0.390	0.417	0.426	1.170	0.434	0.473	0.790
RMSE	120	2	0	0.374	0.376	0.376	0.381	0.408	0.380	**0.371**	0.704
RMSE	120	2	2	**0.371**	0.373	0.377	0.380	0.484	0.385	0.383	0.704
RMSE	120	4	0	**0.380**	0.382	0.387	0.392	0.476	0.400	0.387	0.798
RMSE	120	4	2	**0.379**	0.381	0.388	0.392	0.579	0.405	0.400	0.798
RMSE	360	2	0	0.361	0.361	0.362	0.362	0.366	0.374	**0.360**	0.706
RMSE	360	2	2	**0.359**	0.360	0.362	0.362	0.381	0.375	0.363	0.706
RMSE	360	4	0	0.365	**0.364**	0.367	0.367	0.380	0.390	0.367	0.798
RMSE	360	4	2	**0.364**	0.364	0.368	0.367	0.397	0.392	0.370	0.798
MAE	60	2	0	**0.310**	0.314	0.315	0.325	0.378	0.311	0.313	0.563
MAE	60	2	2	**0.304**	0.307	0.315	0.321	0.531	0.318	0.336	0.559
MAE	60	4	0	**0.312**	0.314	0.329	0.337	0.528	0.336	0.343	0.638
MAE	60	4	2	**0.311**	0.313	0.334	0.341	0.760	0.347	0.377	0.635
MAE	120	2	0	0.299	0.301	0.301	0.305	0.320	0.304	**0.297**	0.565
MAE	120	2	2	**0.297**	0.298	0.301	0.303	0.370	0.308	0.306	0.564
MAE	120	4	0	**0.304**	0.305	0.310	0.314	0.365	0.319	0.309	0.639
MAE	120	4	2	**0.303**	0.304	0.310	0.313	0.427	0.324	0.319	0.638
MAE	360	2	0	0.288	0.288	0.289	0.289	0.291	0.298	**0.287**	0.564
MAE	360	2	2	**0.287**	0.287	0.289	0.289	0.303	0.300	0.290	0.564
MAE	360	4	0	0.291	**0.291**	0.293	0.293	0.303	0.311	0.293	0.638
MAE	360	4	2	0.290	**0.290**	0.293	0.293	0.315	0.313	0.296	0.637

B.2 DMA Models with Google Trends

Table B.3: Static Case with $GT_t = f(\beta_t)$, with $f(\cdot)$ indicating the estimation procedure for Google Trends as outlined in Section 3.3.2. GT-O_{common} labels the DMA model estimated with, noGT-O_{common} labels the DMA model estimated without Google Trends. The ratio is computed as GT-O_{common} / noGT-O_{common}.

error	obs	n_{stat}	n_{noise}	GT-O_{common}	noGT-O_{common}	Ratio
RMSE	60	2	0	0.355	0.385	0.922
RMSE	60	4	0	0.322	0.413	0.780
RMSE	60	2	2	0.357	0.382	0.934
RMSE	60	4	2	0.323	0.416	0.778
RMSE	120	2	0	0.345	0.373	0.926
RMSE	120	4	0	0.317	0.395	0.804
RMSE	120	2	2	0.347	0.370	0.936
RMSE	120	4	2	0.315	0.390	0.807
RMSE	360	2	0	0.336	0.362	0.929
RMSE	360	4	0	0.304	0.372	0.817
RMSE	360	2	2	0.336	0.360	0.933
RMSE	360	4	2	0.304	0.371	0.819
MAE	60	2	0	0.285	0.308	0.923
MAE	60	4	0	0.255	0.330	0.773
MAE	60	2	2	0.285	0.305	0.935
MAE	60	4	2	0.256	0.332	0.771
MAE	120	2	0	0.276	0.298	0.926
MAE	120	4	0	0.252	0.315	0.798
MAE	120	2	2	0.277	0.296	0.935
MAE	120	4	2	0.250	0.312	0.800
MAE	360	2	0	0.269	0.289	0.930
MAE	360	4	0	0.240	0.296	0.810
MAE	360	2	2	0.269	0.288	0.934
MAE	360	4	2	0.240	0.296	0.812

Table B.4: Dynamic Case with $GT_t = f(\beta_t)$. For further notes, see Table B.3.

error	obs	n_{stat}	n_{noise}	GT-O_{common}	noGT-O_{common}	Ratio
RMSE	60	2	0	0.364	0.389	0.937
RMSE	60	4	0	0.323	0.387	0.833
RMSE	60	2	2	0.363	0.383	0.949
RMSE	60	4	2	0.324	0.388	0.834
RMSE	120	2	0	0.350	0.372	0.941
RMSE	120	4	0	0.322	0.380	0.847
RMSE	120	2	2	0.349	0.369	0.947
RMSE	120	4	2	0.323	0.380	0.849
RMSE	360	2	0	0.338	0.361	0.938
RMSE	360	4	0	0.316	0.365	0.867
RMSE	360	2	2	0.339	0.360	0.943
RMSE	360	4	2	0.317	0.365	0.869
MAE	60	2	0	0.291	0.311	0.935
MAE	60	4	0	0.256	0.310	0.826
MAE	60	2	2	0.290	0.307	0.946
MAE	60	4	2	0.257	0.311	0.827
MAE	120	2	0	0.279	0.297	0.940
MAE	120	4	0	0.257	0.304	0.844
MAE	120	2	2	0.279	0.295	0.946
MAE	120	4	2	0.257	0.304	0.846
MAE	360	2	0	0.270	0.288	0.937
MAE	360	4	0	0.252	0.291	0.866
MAE	360	2	2	0.270	0.287	0.943
MAE	360	4	2	0.253	0.291	0.867

Table B.5: Static Case with $GT_t = f(\beta_t) + \epsilon_t$, with $\epsilon_t \sim \mathcal{N}(0, 0.01)$. For further notes, see Table B.3.

error	obs	n_{stat}	n_{noise}	GT-O_{common}	noGT-O_{common}	Ratio
RMSE	60	2	0	0.358	0.386	0.926
RMSE	60	4	0	0.322	0.413	0.779
RMSE	60	2	2	0.359	0.384	0.934
RMSE	60	4	2	0.323	0.412	0.783
RMSE	120	2	0	0.348	0.374	0.931
RMSE	120	4	0	0.317	0.393	0.807
RMSE	120	2	2	0.345	0.369	0.934
RMSE	120	4	2	0.315	0.390	0.807
RMSE	360	2	0	0.339	0.363	0.932
RMSE	360	4	0	0.305	0.371	0.822
RMSE	360	2	2	0.338	0.362	0.933
RMSE	360	4	2	0.305	0.372	0.822
MAE	60	2	0	0.287	0.309	0.928
MAE	60	4	0	0.255	0.330	0.774
MAE	60	2	2	0.287	0.307	0.935
MAE	60	4	2	0.257	0.330	0.778
MAE	120	2	0	0.279	0.299	0.933
MAE	120	4	0	0.252	0.314	0.803
MAE	120	2	2	0.276	0.296	0.934
MAE	120	4	2	0.250	0.312	0.803
MAE	360	2	0	0.271	0.290	0.935
MAE	360	4	0	0.242	0.296	0.816
MAE	360	2	2	0.270	0.289	0.934
MAE	360	4	2	0.242	0.297	0.815

Table B.6: Dynamic Case with $GT_t = f(\beta_t) + \epsilon_t$, with $\epsilon_t \sim \mathcal{N}(0, 0.01)$. For further notes, see Table B.3.

error	obs	n_{stat}	n_{noise}	GT-O_{common}	noGT-O_{common}	Ratio
RMSE	60	2	0	0.364	0.389	0.936
RMSE	60	4	0	0.324	0.387	0.838
RMSE	60	2	2	0.362	0.382	0.945
RMSE	60	4	2	0.324	0.386	0.840
RMSE	120	2	0	0.352	0.372	0.945
RMSE	120	4	0	0.325	0.381	0.853
RMSE	120	2	2	0.352	0.372	0.948
RMSE	120	4	2	0.323	0.380	0.850
RMSE	360	2	0	0.339	0.361	0.940
RMSE	360	4	0	0.317	0.365	0.870
RMSE	360	2	2	0.339	0.360	0.942
RMSE	360	4	2	0.316	0.364	0.868
MAE	60	2	0	0.292	0.312	0.935
MAE	60	4	0	0.258	0.310	0.834
MAE	60	2	2	0.289	0.306	0.943
MAE	60	4	2	0.258	0.310	0.833
MAE	120	2	0	0.281	0.297	0.944
MAE	120	4	0	0.259	0.305	0.850
MAE	120	2	2	0.281	0.297	0.947
MAE	120	4	2	0.257	0.304	0.847
MAE	360	2	0	0.271	0.288	0.941
MAE	360	4	0	0.253	0.291	0.868
MAE	360	2	2	0.270	0.287	0.942
MAE	360	4	2	0.252	0.291	0.866

Table B.7: Static Case with $GT_t = f(\beta_t) + \epsilon_t$, with $\epsilon_t \sim \mathcal{N}(0, 0.1)$. For further notes, see Table B.3.

error	obs	n_{stat}	n_{noise}	GT-O_{common}	noGT-O_{common}	Ratio
RMSE	60	2	0	0.367	0.384	0.955
RMSE	60	4	0	0.355	0.413	0.860
RMSE	60	2	2	0.358	0.382	0.938
RMSE	60	4	2	0.350	0.415	0.844
RMSE	120	2	0	0.357	0.373	0.956
RMSE	120	4	0	0.345	0.393	0.876
RMSE	120	2	2	0.349	0.371	0.941
RMSE	120	4	2	0.337	0.392	0.859
RMSE	360	2	0	0.346	0.362	0.957
RMSE	360	4	0	0.328	0.371	0.885
RMSE	360	2	2	0.339	0.361	0.940
RMSE	360	4	2	0.323	0.371	0.872
MAE	60	2	0	0.294	0.307	0.956
MAE	60	4	0	0.283	0.330	0.857
MAE	60	2	2	0.286	0.306	0.936
MAE	60	4	2	0.277	0.331	0.838
MAE	120	2	0	0.285	0.298	0.956
MAE	120	4	0	0.274	0.314	0.875
MAE	120	2	2	0.278	0.296	0.939
MAE	120	4	2	0.268	0.314	0.852
MAE	360	2	0	0.277	0.289	0.957
MAE	360	4	0	0.261	0.296	0.883
MAE	360	2	2	0.270	0.288	0.937
MAE	360	4	2	0.255	0.296	0.864

Table B.8: Dynamic Case with $GT_t = f(\beta_t) + \epsilon_t$, with $\epsilon_t \sim \mathcal{N}(0, 0.1)$. For further notes, see Table B.3.

error	obs	n_{stat}	n_{noise}	GT-O_{common}	noGT-O_{common}	Ratio
RMSE	60	2	0	0.373	0.387	0.963
RMSE	60	4	0	0.352	0.389	0.905
RMSE	60	2	2	0.363	0.383	0.949
RMSE	60	4	2	0.343	0.384	0.894
RMSE	120	2	0	0.359	0.372	0.965
RMSE	120	4	0	0.346	0.380	0.910
RMSE	120	2	2	0.353	0.371	0.953
RMSE	120	4	2	0.339	0.380	0.892
RMSE	360	2	0	0.347	0.361	0.962
RMSE	360	4	0	0.335	0.365	0.917
RMSE	360	2	2	0.341	0.359	0.949
RMSE	360	4	2	0.329	0.365	0.902
MAE	60	2	0	0.299	0.311	0.962
MAE	60	4	0	0.281	0.311	0.901
MAE	60	2	2	0.290	0.306	0.946
MAE	60	4	2	0.272	0.307	0.888
MAE	120	2	0	0.287	0.298	0.963
MAE	120	4	0	0.276	0.304	0.909
MAE	120	2	2	0.282	0.297	0.950
MAE	120	4	2	0.270	0.304	0.889
MAE	360	2	0	0.277	0.288	0.962
MAE	360	4	0	0.267	0.291	0.917
MAE	360	2	2	0.271	0.287	0.947
MAE	360	4	2	0.262	0.292	0.899

Table B.9: Static Case with $GT_t = \epsilon_t$, with $\epsilon_t \sim \mathcal{U}(0,1)$.

error	obs	n_{stat}	n_{noise}	GT-O_{common}	noGT-O_{common}	Ratio
RMSE	60	2	0	0.418	0.385	1.085
RMSE	60	4	0	0.483	0.416	1.163
RMSE	60	2	2	0.428	0.380	1.129
RMSE	60	4	2	0.496	0.422	1.177
RMSE	120	2	0	0.414	0.372	1.113
RMSE	120	4	0	0.461	0.395	1.168
RMSE	120	2	2	0.414	0.365	1.133
RMSE	120	4	2	0.477	0.394	1.210
RMSE	360	2	0	0.405	0.360	1.124
RMSE	360	4	0	0.441	0.369	1.193
RMSE	360	2	2	0.408	0.357	1.142
RMSE	360	4	2	0.455	0.370	1.228
MAE	60	2	0	0.332	0.308	1.077
MAE	60	4	0	0.384	0.332	1.156
MAE	60	2	2	0.340	0.306	1.109
MAE	60	4	2	0.394	0.337	1.171
MAE	120	2	0	0.328	0.297	1.104
MAE	120	4	0	0.365	0.316	1.158
MAE	120	2	2	0.329	0.293	1.121
MAE	120	4	2	0.379	0.315	1.204
MAE	360	2	0	0.320	0.288	1.110
MAE	360	4	0	0.347	0.295	1.179
MAE	360	2	2	0.322	0.285	1.130
MAE	360	4	2	0.359	0.296	1.213

Table B.10: Dynamic Case with $GT_t = \epsilon_t$, with $\epsilon_t \sim \mathcal{U}(0,1)$.

error	obs	n_{stat}	n_{noise}	GT-O_{common}	noGT-O_{common}	Ratio
RMSE	60	2	0	0.412	0.385	1.069
RMSE	60	4	0	0.443	0.389	1.140
RMSE	60	2	2	0.427	0.389	1.099
RMSE	60	4	2	0.447	0.378	1.182
RMSE	120	2	0	0.410	0.368	1.115
RMSE	120	4	0	0.428	0.384	1.115
RMSE	120	2	2	0.417	0.372	1.121
RMSE	120	4	2	0.432	0.376	1.151
RMSE	360	2	0	0.403	0.361	1.118
RMSE	360	4	0	0.413	0.365	1.130
RMSE	360	2	2	0.409	0.359	1.138
RMSE	360	4	2	0.423	0.367	1.151
MAE	60	2	0	0.324	0.309	1.048
MAE	60	4	0	0.341	0.309	1.104
MAE	60	2	2	0.340	0.311	1.091
MAE	60	4	2	0.355	0.301	1.180
MAE	120	2	0	0.313	0.299	1.048
MAE	120	4	0	0.322	0.303	1.064
MAE	120	2	2	0.332	0.298	1.116
MAE	120	4	2	0.346	0.301	1.150
MAE	360	2	0	0.304	0.288	1.054
MAE	360	4	0	0.309	0.292	1.056
MAE	360	2	2	0.324	0.287	1.130
MAE	360	4	2	0.335	0.294	1.141

B.3 Empirical Application

B.3.1 Data

Economic Activity and Demand (EAAD)

Macroeconomic variable: Industrial Production (IP)
Google Trends: industrial production, supply bottlenecks, raw material prices, production costs, construction projects, construction companies, automotive industry

Macroeconomic variable: Retail Trade Sales (RT)
Google Trends: retail trade sales, online shopping, retail, discount campaigns, shop opening hours, discounts, black friday

Macroeconomic variable: Household Consumption (HC)
Google Trends: household consumption, household expenditure, consumer behavior, saving measures, budget planning, cost of living, energy costs

Macroeconomic variable: Price-to-income-Ratio (PIR)
Google Trends: property bubble, rent price, salary increase, inflation, salary

Macroeconomic variable: Savings Rate (SR)
Google Trends: savings rate, savings interest, financial planning, retirement provision, interest rates

Macroeconomic variable: Further EAAD-terms
Google Trends: economic crisis, traveling expenses, property market, tourism expenditure, economy, tourism, building a house, residential investment, recession, holiday apartment, house prices, economic growth, plane ticket, house construction, new building, hotel costs, renovation costs

Labor Market and Employment (LMAE)

Macroeconomic variable: Unemployment Rate (UR)
Google Trends: unemployment rate, labor office, job loss, unemployment benefit, long-term unemployed, economic crisis

Macroeconomic variable: Part-time Employment Rate (PT)
Google Trends: minijobs, part-time jobs, part time

Macroeconomic variable: Short-time Employment Rate (STW)
Google Trends: reduction in working hours, short-time working scheme, economic aid

Macroeconomic variable: Short-time Volume Loss (STL)
Google Trends: employment, job cuts

Macroeconomic variable: Labor Volume (LV)
Google Trends: labor volume, labor market, employment

Macroeconomic variable: Multiple Job Holders (MJ)
Google Trends: multiple employment, additional income, second job, freelancer, self-employed

Macroeconomic variable: Job Vacancy Rate (JV)
Google Trends: vacancies, open positions, jobs, personnel, job portal

Macroeconomic variable: Employment Growth (EG)
Google Trends: new hires, labor market

Macroeconomic variable: Further LMAE-terms
Google Trends: employment agency, skilled labor shortage, youth unemployment rate, register unemployed online, working time models, education, long-term unemployment rate, employment rate, job center application, working life, internship, short-time work, salary, social benefits, job exchange, labor office, long-term unemployment, vocational training, unemployed

Trade and External Sector (TAES)

Macroeconomic variable: Ifo Export Expectations (IE)
Google Trends: exports, world economy, trade balance

Macroeconomic variable: Manufacturing PMI (PMI)
Google Trends: purchasing managers' index, industrial production, industry, supply chain

Macroeconomic variable: Baltic Dry Index (BDI)
Google Trends: freight rates, trade, container, commodity prices, transport costs, sea freight market

Macroeconomic variable: Bundesbank Lending Indicator (LI)
Google Trends: lending, interest rates, corporate financing, debt capital, monetary policy, investments

Macroeconomic variable: Net Export Growth (NE)
Google Trends: import export, trade surplus

Macroeconomic variable: Real Effective Exchange Rate (REER)
Google Trends: exchange rate, euro, dollar, monetary policy

Macroeconomic variable: Further TAES-terms
Google Trends: customs duties, tariffs, exports, made in germany, brexit, balance of trade, vw diesel scandal, foreign trade, free trade agreement, currency development

Financial Market Conditions (FMC)

Macroeconomic variable: DAX Index (DAX)
Google Trends: stock market, german shares, dax forecast, stock market trading, volatility

Macroeconomic variable: Volatility Index (VDAX)
Google Trends: market risk, uncertainty indicator

Macroeconomic variable: Money Supply Growth (M3)
Google Trends: money supply, central bank, inflation, interest rate

Macroeconomic variable: Bank Lending Standards (BLS)
Google Trends: bank loans, equity capital, corporate financing

Macroeconomic variable: Term Spread (TS)
Google Trends: yield curve, phillips curve, bonds, yield

Macroeconomic variable: Credit Spread (CS)
Google Trends: credit risk, bonds, credit rating, corporate bonds, risk premium

Macroeconomic variable: Further FMC-terms
Google Trends: exchange rate, government debt, consumer credit growth, key interest rate, interest rate development, government bond, interest rates, financial market, ecb monetary policy, fed, credit, vdax, prices

Price Level and Inflation (PLAI)

Macroeconomic variable: CPI Inflation (CPI)
Google Trends: consumer price index, inflation rate, price increases, consumer prices

Macroeconomic variable: Purchasing Price Inflation (PPI)
Google Trends: producer price index, production costs, raw material prices, cost increases

Macroeconomic variable: Oil Price Inflation (OPI)
Google Trends: oil prices, petrol prices, crude oil prices

Macroeconomic variable: Energy Price Inflation (EPI)
Google Trends: electricity prices, heating costs, energy crisis, gas prices, energy consumption, energy prices

Macroeconomic variable: Wage Inflation (WI)
Google Trends: wage increase, wage development, collective bargaining, minimum wage, labor costs

Macroeconomic variable: Further PLAI-terms
Google Trends: price development, prices, production costs, cost of living, inflation, IG Metall

Sentiments and Expectations (SAE)

Macroeconomic variable: Ifo Business Climate (IFO)
Google Trends: business climate index, clemens fuest, business cycle, economy, ifo index

Macroeconomic variable: GfK Consumer Confidence (GFK)
Google Trends: consumer climate

Macroeconomic variable: ZEW Economic Sentiment (ZEW)
Google Trends: zew mannheim

Macroeconomic variable: Further SAE-terms
Google Trends: ifo index, financial market, stock market

Real Estate and Housing (REAH)

Macroeconomic variable: Building Permits (BP)
Google Trends: new building project, building applications, building activity, construction sector

Macroeconomic variable: Residential Property Prices (RPP)
Google Trends: property prices, house prices, property market

Macroeconomic variable: Construction Activity (CA)
Google Trends: construction projects, construction industry

Macroeconomic variable: Residential Construction (RC)
Google Trends: housing construction, property, buying a house

Macroeconomic variable: Further REAH-terms
Google Trends: rents, building industry, property crisis, property bubble

B.3.2 Forecasting GDP

Evolution of forecasted GDP growth rates (Section 3.4.2)

Figure B.1: Forecasting y_t: Forecasted Values

B.3.3 Identifying Variables Overlooked in Expert GDP Forecasts

Forward Selection Algorithm (Section 3.4.3)

Algorithm 1 Forward Selection DMA Estimation

Initialize the set of explanatory variables $X = \{x_1, x_2, \ldots, x_{37}\}$
Set the maximum number of variables $N = 10$
Initialize the selected variables set $S = \{\}$
Initialize the best fit value $MAE_{best} = \infty$
for each variable $x_i \in X$ **do**
    Estimate $\hat{y} = f_{DMA}(x_i)$
    Compute MAE based on one-step-ahead forecasts
    if $MAE < MAE_{best}$ **then**
        $MAE_{best} \leftarrow$ fit
        $S \leftarrow \{x_i\}$
    end if
end for
while size of $S < N$ **do**
    Let $X_{rest} \leftarrow X \backslash S$
    for each variable $x_j \in X_{rest}$ **do**
        Estimate $\hat{y} = f_{DMA}(S \cup \{x_j\})$
        Compute MAE based on one-step-ahead forecasts
        if new $MAE < MAE_{best}$ **then**
            $MAE_{best} \leftarrow MAE$
            Add x_j to S
        end if
    end for
end while
return S

Interpretation of Marginal Effects (Section 3.4.3)

As outlined in Section 3.4.3, we assume that the experts consider the same macroeconomic quantities as we do when they build their forecasts. Accordingly, if we define the difference between true GDP growth and the (mean) expert forecast as the dependent variable and regress this difference on the macroeconomic variables, significances indicate which variables were not sufficiently considered in the formulation of the experts' forecast. More formally, the expert's forecast error $\tilde{y}_t$ can be expressed as $\tilde{y}_t = \sum_{k=1}^{K} (\partial y_t / \partial X_{k,t} - \partial f(X_{k,t}) / \partial X_{k,t})$ with K representing the number of columns in $X_{k,t}$. Accordingly, if

$\frac{\partial \tilde{y}_t}{\partial X_t} > 0$: experts underweight the information contained in X_t in their forecasts as

$$\partial f(X_t)/\partial X_t < \partial y/\partial X_t$$

$\frac{\partial \tilde{y}_t}{\partial X_t} < 0$: experts overweight the information contained in X_t in their forecasts as

$$\partial f(X_t)/\partial X_t > \partial y_t/\partial X_t$$

$\frac{\partial \tilde{y}_t}{\partial X_t} = 0$: experts accurately incorporate the information from X_t into their forecasts as

$$\partial f(X_t)/\partial X_t = \partial y_t/\partial X_t$$

Evolution of forecasted GDP growth rates (Section 3.4.3)

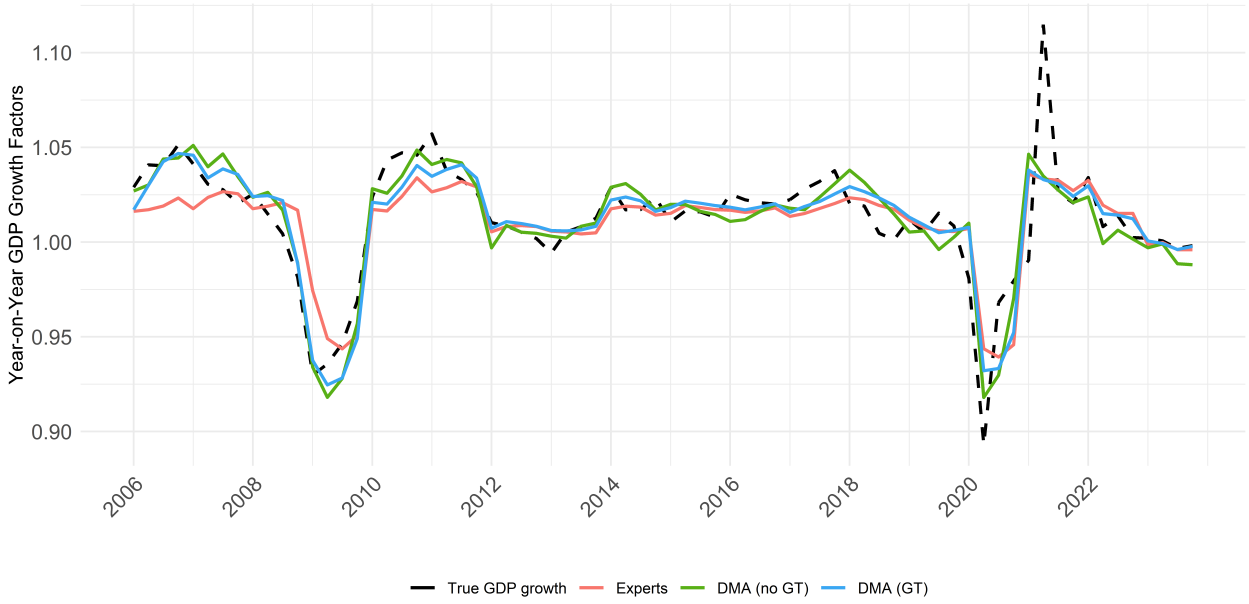

Figure B.2: Forecasting $\tilde{y}_t$: Forecasted Values

Evolution of overlooked coefficients (Section 3.4.3)

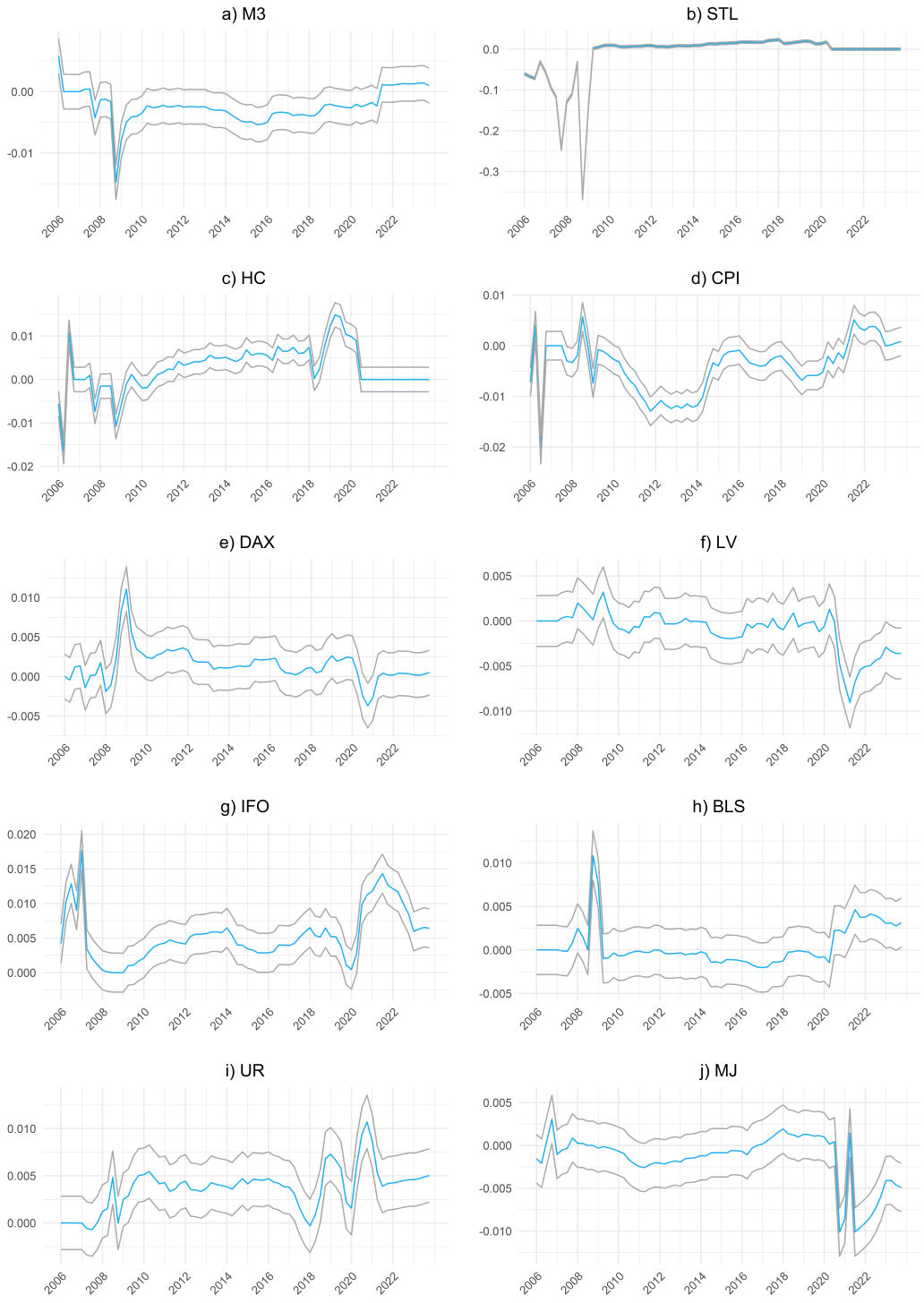

Figure B.3: Estimated dynamic coefficients of the 10 selected macroeconomic variables using DMA with $\tilde{y}_t$ as dependent variable and no Google Trends. Coefficients are blue, confidence bands are grey.

C Appendix for Chapter 4

C.1 Homogeneous Parameter Estimation

It is useful to briefly summarize how the homogeneous parameters ρ, β and γ as well as the variances components σ_λ^2 and σ_ε^2 can be estimated in a standard dynamic panel. The procedures below are well known and are presented for the sake of completeness.

Quasi–Maximum Likelihood

Under the working assumption of Gaussian innovations, the unit effects can be integrated out to obtain a likelihood depending only on the homogeneous parameters. Maximizing this marginal likelihood yields the *quasi*–maximum likelihood estimator (QMLE) of ρ and β. It is consistent under mild regularity conditions and typically remains reliable even if the exact normality assumption is violated (see, e.g., Moral-Benito et al., 2019).

Generalized Method of Moments

Moment–based estimators exploit orthogonality conditions between lagged variables and the innovation. The *Difference GMM* (Arellano and Bond, 1991) first–differences the model to remove the unit effect and uses deeper lags as instruments for the lagged dependent variable. The *Level GMM* (Arellano and Bover, 1995) replaces first differences by forward orthogonal deviations, which retain more information and alleviate weak instrument problems when the series is highly persistent. The *System GMM* (Blundell and Bond, 1998) combines the difference and level equations in a single system, using lagged differences as instruments for the level equation to improve efficiency when persistence is high. These GMM estimators are designed for panels with many individuals and few time periods and deliver consistent estimates of ρ and β.

GLS Estimation with Correlated Random Effects

Including the time–average $\bar{x}_i$ ensures that the remaining individual component η_i is orthogonal to the regressors, yielding the residuals

$$r_{it} = \begin{cases} v_{it} = y_{it} - \bar{x}_i\gamma = \rho\, y_{i,t-1} + \eta_i + \varepsilon_{it} & \text{under stationary IC,} \\ z_{it} = y_{it} - \rho\, y_{i,t-1} - \bar{x}_i\gamma = \eta_i + \varepsilon_{it} & \text{under non-stationary IC.} \end{cases}$$

Because η_i enters each time period, the composite error term exhibits a variance components structure with serial correlation within units. Ordinary least squares ignores this dependence and is therefore inefficient. A generalized least squares (GLS) estimator accounts for the covariance structure implied by the repeated effect η_i and provides efficient estimates of γ (see, e.g., Wooldridge, 2010).

Variance–component estimation

When a random–effects specification is adopted, the variance components $(\sigma^2_\lambda, \sigma^2_\varepsilon)$ can be estimated by the classical Swamy–Arora ANOVA–type procedure (Swamy and Arora, 1972). The method decomposes the variance of the process into a within–individual component, yielding an estimate of the idiosyncratic variance σ^2_ε, and a between–individual component, from which σ^2_λ is recovered after subtracting the sampling noise $\sigma^2_\varepsilon/(T-1)$. For simplicity, ignore additional explanatory variables and consider the series under non-stationary IC:

$$\begin{aligned} z_{it} &= y_{it} - \rho\, y_{i,t-1} \\ &= \lambda_i + \varepsilon_{it}, \qquad t = 2, \ldots, T. \end{aligned}$$

Define the time average $\bar{z}_i := \frac{1}{T-1}\sum_{t=2}^{T} z_{it}$. The total variability of $\{z_{it}\}$ can be decomposed into a within–unit and a between–unit component. The *within* mean–square provides a consistent estimator of the innovation variance:

$$\hat{\sigma}^2_\varepsilon = \frac{1}{N(T-2)} \sum_{i=1}^{N} \sum_{t=2}^{T} \big(z_{it} - \bar{z}_i\big)^2.$$

The dispersion of $\bar{z}_i$ reflects both the cross–sectional variation in λ_i and the sampling noise from averaging over $T-1$ observations. Removing the latter yields the Swamy–Arora estimator of the variance of the unit effects:

$$\hat{\sigma}^2_\lambda = \mathrm{Var}(\bar{z}_i) - \frac{\hat{\sigma}^2_\varepsilon}{T-1}.$$

C.2 Autocovariances under Stationary Conditions

Under stationary conditions, the transformed series relevant for forecasting is

$$y_{it} = \rho\, y_{i,t-1} + \lambda_i + \varepsilon_{it},$$

where λ_i is the time-invariant unit effect with variance σ_λ^2 and ε_{it} is an innovation with mean zero, variance σ_ε^2, and no serial correlation. Because λ_i enters every period and y_{it} follows a stationary AR(1) process around the unit mean, y_{it} inherits two sources of persistence: the lag dependence governed by ρ and the constant unit-specific component.
To derive the autocovariances of y_{it}, it is useful to recall the stationary AR(1) component of the model. Let $u_{it} := y_{it} - \mu_i$ denote the demeaned series. Then

$$u_{it} = \rho\, u_{i,t-1} + \varepsilon_{it}.$$

Taking the variances on both sides and noting that u_{it} and ε_{it} are uncorrelated by construction,

$$\mathrm{Var}(u_{it}) = \mathrm{Var}(\rho\, u_{i,t-1} + \varepsilon_{it}) = \rho^2\, \mathrm{Var}(u_{it}) + \sigma_\varepsilon^2.$$

Solving for $\mathrm{Var}(u_{it})$ yields the well-known stationary AR(1) result,

$$\mathrm{Var}(u_{it}) = \frac{\sigma_\varepsilon^2}{1-\rho^2}.$$

Because y_{it} contains both λ_i and the AR(1) component, its variance is

$$\gamma_0^{(y)} := \mathrm{Var}(y_{it}) = \sigma_\lambda^2 + \frac{\sigma_\varepsilon^2}{1-\rho^2}.$$

For the lag-k autocovariance, note that λ_i contributes σ_λ^2 at every lag, while the AR(1) component decays geometrically at rate ρ^k. Hence

$$\gamma_k^{(y)} := \mathrm{Cov}(y_{it}, y_{i,t-k}) = \sigma_\lambda^2 + \rho^k\, \frac{\sigma_\varepsilon^2}{1-\rho^2}, \qquad k = 1, 2, \ldots.$$

The first term generates persistence that does not decay with k, while the second term captures dynamic dependence that vanishes as k grows. This autocovariance structure determines the coefficients of the linear one-step-ahead predictor used in Section 4.2.3.

C.3 Autocovariances under Non-Stationary Conditions

Under non-stationary conditions, the transformed series relevant for forecasting is the ρ-residual process

$$z_{it} = \lambda_i + \varepsilon_{it}, \qquad t = 2, \ldots, T.$$

To derive the posterior mean of λ_i given the transformed history, we look for the MSE-optimal linear predictor of the form

$$E[\lambda_i \mid z_{i2}, \ldots, z_{iT}] = \sum_{t=2}^{T} b_t z_{it}.$$

Note that each transformed observation satisfies $z_{it} = \lambda_i + \varepsilon_{it}$ with the same signal part λ_i and i.i.d. noise. Therefore, the optimal coefficients must all be equal and the MSE-optimal predictor reduces to

$$E[\lambda_i \mid z_{i2}, \ldots, z_{iT}] = \tilde{b}\, \bar{z}_i, \qquad \tilde{b} = (T-1)\, b,$$

where

$$\bar{z}_i := \frac{1}{T-1} \sum_{t=2}^{T} z_{it}.$$

Because

$$\mathbb{E}[\bar{z}_i \mid \lambda_i] = \lambda_i, \qquad \mathrm{Var}(\bar{z}_i) = \mathrm{Var}(\lambda_i) + \mathrm{Var}\left(\frac{1}{T-1} \sum_{t=2}^{T} \varepsilon_{it} \right) = \sigma_\lambda^2 + \frac{\sigma_\varepsilon^2}{T-1},$$

and

$$\mathrm{Cov}(\lambda_i, \bar{z}_i) = \sigma_\lambda^2,$$

the linear projection of λ_i on the transformed history $\{z_{i2}, \ldots, z_{iT}\}$ reduces to a reliability-weighted mean:

$$\mathbb{E}[\lambda_i \mid z_{i2}, \ldots, z_{iT}] = \frac{\mathrm{Cov}(\lambda_i, \bar{z}_i)}{\mathrm{Var}(\bar{z}_i)}\, \bar{z}_i = \kappa\, \bar{z}_i, \qquad \kappa := \frac{\sigma_\lambda^2}{\sigma_\lambda^2 + \sigma_\varepsilon^2/(T-1)}.$$

Substituting this predictor gives the MSE–optimal one-step forecast under non-stationary IC:

$$\widehat{y}_{i,T+1|T} = \rho\, y_{iT} + \kappa\, \bar{z}_i.$$

C.4 Empirical Application

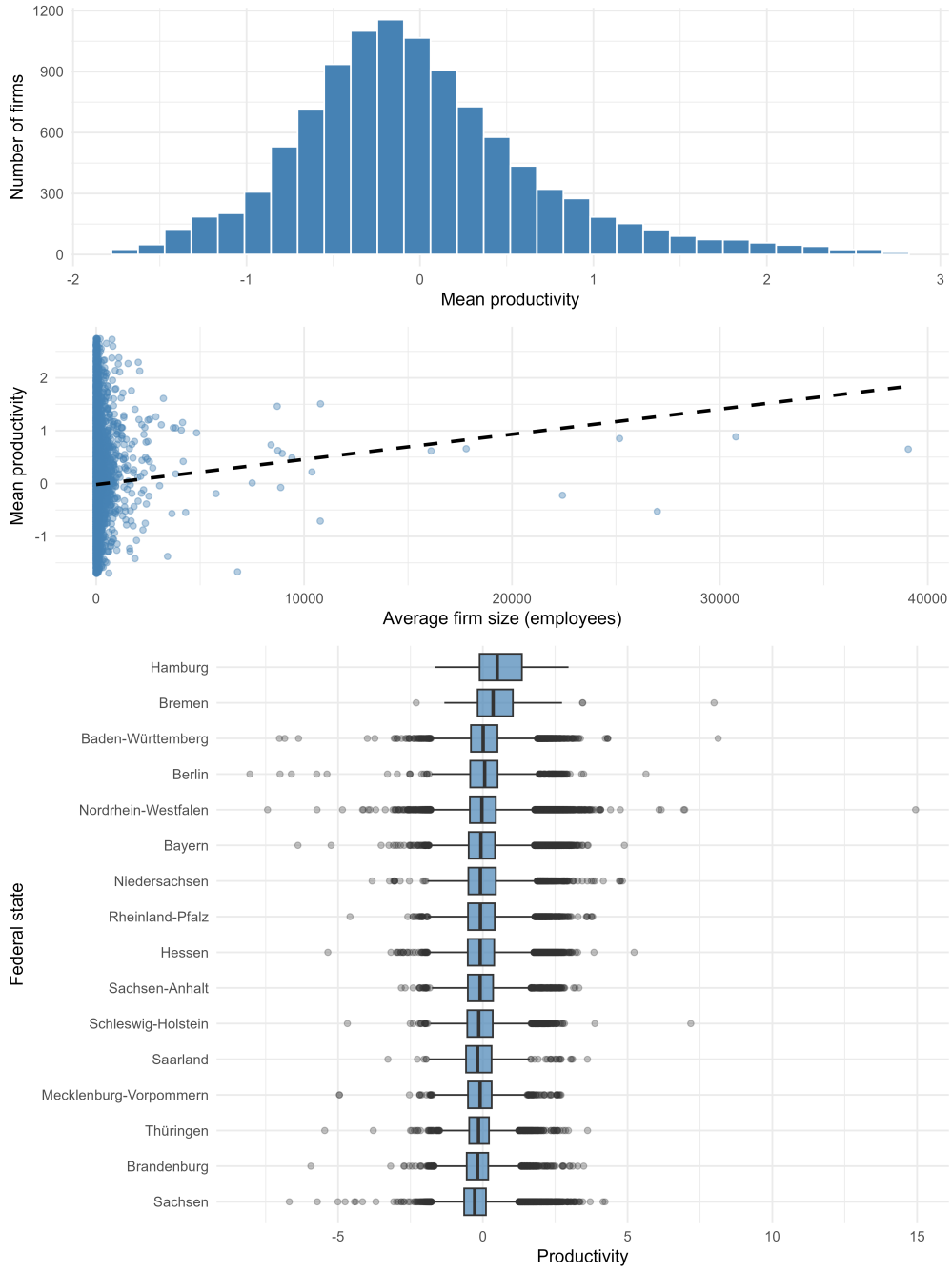

Figure C.1: Firm-Level Productivity and Observable Characteristics. The upper panel shows the distribution of mean productivity across firms. The middle panel depicts the relationship between average firm size and mean productivity, with a fitted linear trend. The lower panel shows differences in productivity, sorted across federal states.

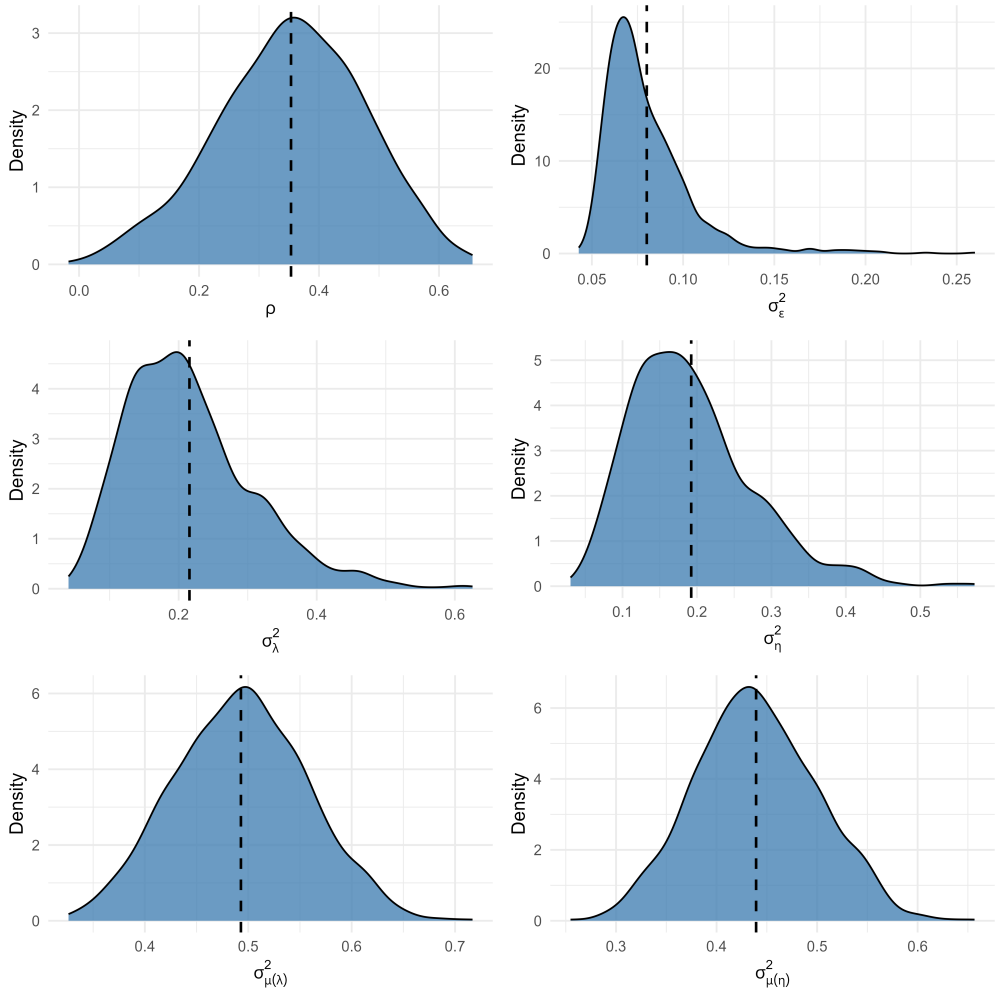

Figure C.2: Common Coefficient Densities for $T = 12$

Bibliography

Abadie, A., A. Diamond, and J. Hainmueller (2010, 02). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association 105*, 493–505.

Abadie, A., A. Diamond, and J. Hainmueller (2015). Comparative politics and the synthetic control method. *American Journal of Political Science 59*(2), 495–510.

Abadie, A. and J. Gardeazabal (2003, 02). The economic costs of conflict: A case study of the basque country. *American Economic Review 93*, 113–132.

Arel-Bundock, V. (2022). *WDI: World Development Indicators and Other World Bank Data*. R package version 2.7.8.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies 58*(2), 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics 68*(1), 29–51.

Athey, S. and G. Imbens (2016, 07). The state of applied econometrics - causality and policy evaluation. *Journal of Economic Perspectives 31*.

Bahrami, S., K. Hajian-Tilaki, M. Bayani, M. Chehrazi, Z. Mohamadi-Pirouz, and A. Amoozadeh (2023). Bayesian model averaging for predicting factors associated with length of covid-19 hospitalization. *BMC Medical Research Methodology 23*(1), 163.

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

Baltagi, B. H. (2008). Forecasting with panel data. *Journal of forecasting 27*(2), 153–173.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large bayesian vector auto regressions. *Journal of applied Econometrics 25*(1), 71–92.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica: Journal of the Econometric Society*, 657–681.

Belmonte, M. and G. Koop (2014). Model switching and model averaging in time-varying parameter regression models. In *Bayesian Model Comparison*, Volume 34, pp. 45–69. Emerald Group Publishing Limited.

Ben-Michael, E., A. Feller, and J. Rothstein (2021, 01). The augmented synthetic control method. *SSRN Electronic Journal*.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics 87*(1), 115–143.

Bork, L. and S. V. Møller (2015). Forecasting house prices in the 50 states using dynamic model averaging and dynamic model selection. *International Journal of Forecasting 31*(1), 63–78.

Catania, L. and N. Nonejad (2016). Dynamic model averaging for practitioners in economics and finance: The edma package. *arXiv preprint arXiv:1606.05656*.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of econometrics 18*(1), 5–46.

Chen, B. and K. Maung (2023). Time-varying forecast combination for high-dimensional data. *Journal of Econometrics 237*(2, Part C), 105418.

Clements, M. P., R. W. Rich, and J. S. Tracy (2023). Surveys of professionals. In *Handbook of economic expectations*, pp. 71–106. Elsevier.

de Jong, S. and H. A. Kiers (1992). Principal covariates regression: Part i. theory. *Chemometrics and Intelligent Laboratory Systems 14*(1), 155–164. Proceedings of the 2nd Scandinavian Symposium on Chemometrics.

Doudchenko, N. and G. W. Imbens (2016, October). Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis. NBER Working Papers 22791, National Bureau of Economic Research, Inc.

Drachal, K. (2016). Forecasting spot oil price in a dynamic model averaging framework — have the determinants changed over time? *Energy Economics 60*, 35–46.

Drachal, K. (2020). Dynamic model averaging in economics and finance with fdma: A package for r. *Signals 1*(1), 4.

Efron, B. (2011). Tweedie's formula and selection bias. *Journal of the American Statistical Association 106*(496), 1602–1614.

Fang, M. and X. Li (2016). Application of bayesian model averaging in the reconstruction of past climate change using pmip3/cmip5 multimodel ensemble simulations. *Journal of climate 29*(1), 175–189.

Ferman, B. (2021, October). On the Properties of the Synthetic Control Estimator with Many Periods and Many Controls. *Journal of the American Statistical Association 116*(536), 1764–1772.

Galvão, A. B., A. Garratt, and J. Mitchell (2021). Does judgment improve macroeconomic density forecasts? *International Journal of Forecasting 37*(3), 1247–1260.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association 57*(298), 369–375.

Hall, P., J. L. Horowitz, and B.-Y. Jing (1995). On blocking rules for the bootstrap with dependent data. *Biometrika 82*(3), 561–574.

Harvey, A. and S. Thiele (2020, 12). Cointegration and control: Assessing the impact of events using time series data. *Journal of Applied Econometrics 36*.

Hendry, D. F. and M. P. Clements (2004, 06). Pooling of forecasts. *The Econometrics Journal 7*(1), 1–31.

Hoerl, A. E. and R. W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics 12*, 55–67.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association 81*(396), 945–960.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.

Koop, G. and D. Korobilis (2012). Forecasting inflation using dynamic model averaging. *International Economic Review 53*(3), 867–886.

Koop, G. and L. Onorante (2019). Macroeconomic nowcasting using google probabilities. In *Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling: Part A*, pp. 17–40. Emerald Publishing Limited.

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

Li, J. and Y. Jiang (2022). Recent advances of dynamic model averaging theory and its application in econometrics. *Journal of Financial Risk Management 11*(4), 740–756.

Liu, L., H. R. Moon, and F. Schorfheide (2020). Forecasting with dynamic panel data models. *Econometrica 88*(1), 171–201.

McCormick, T. H., A. E. Raftery, D. Madigan, and R. S. Burd (2012). Dynamic logistic regression and dynamic model averaging for binary classification. *Biometrics 68*(1), 23–30.

Montgomery, J. M. and B. Nyhan (2010). Bayesian model averaging: Theoretical developments and practical applications. *Political Analysis 18*(2), 245–270.

Moral-Benito, E., P. Allison, and R. Williams (2019). Dynamic panel data modelling using maximum likelihood: an alternative to arellano-bond. *Applied Economics 51*(20), 2221–2232.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society*, 69–85.

Naser, H. (2016). Estimating and forecasting the real prices of crude oil: A data rich model using a dynamic model averaging (dma) approach. *Energy Economics 56*, 75–87.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science 5 4*, 465–472.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the econometric society*, 1417–1426.

Nonejad, N. (2021). An overview of dynamic model averaging techniques in time-series econometrics. *Journal of Economic Surveys 35*(2), 566–614.

Onorante, L. and A. E. Raftery (2016). Dynamic model averaging in large model spaces using dynamic occams window. *European Economic Review 81*, 2–14.

Ottaviani, M. and P. N. Sørensen (2006). The strategy of professional forecasting. *Journal of Financial Economics 81*(2), 441–466.

Pesaran, H. M. and R. P. Smith (2014, October). Tests of Policy Ineffectiveness in Macroeconometrics. CESifo Working Paper Series 4871.

Pesaran, M. H., A. Pick, and A. Timmermann (2024). Forecasting with panel data: Estimation uncertainty versus parameter heterogeneity. *arXiv preprint arXiv:2404.11198*.

Raftery, A. E., M. Kárnỳ, and P. Ettler (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics 52*(1), 52–66.

Risse, M. and L. Ohl (2017). Using dynamic model averaging in state space representation with dynamic occams window and applications to the stock and gold market. *Journal of Empirical Finance 44*, 158–176.

Robbins, H. (1985). Asymptotically subminimax solutions of compound statistical decision problems. *Herbert Robbins Selected Papers*, 7–24.

Roehl, K.-H. (2025). 35 jahre wiedervereinigung: Was wurde in ostdeutschland erreicht und wo liegen die groessten oekonomischen probleme? *IW-Trends-Vierteljahresschrift zur empirischen Wirtschaftsforschung 52*(3), 61–80.

Rossi, B. (2021). Forecasting in the presence of instabilities: How we know whether models predict well and how to improve them. *Journal of Economic Literature 59*(4), 1135–1190.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*(5), 688–701.

Sala-i Martin, X., G. Doppelhofer, and R. I. Miller (2004). Determinants of long-term growth: A bayesian averaging of classical estimates (bace) approach. *American economic review 94*(4), 813–835.

Scott, S. L. and H. R. Varian (2013). Bayesian variable selection for nowcasting economic time series. (19567).

Scott, S. L. and H. R. Varian (2014). Predicting the present with bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation 5*(1-2), 4–23.

Smets, F., A. Warne, and R. Wouters (2014). Professional forecasters and real-time forecasting with a DSGE model. *International Journal of Forecasting 30*(4), 981–995.

Steel, M. F. (2020). Model averaging and its use in economics. *Journal of Economic Literature 58*(3), 644–719.

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

Stock, J. H. and M. W. Watson (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics*, Volume 2, pp. 415–525. Elsevier.

Swamy, P. A. V. B. and S. S. Arora (1972). The exact finite sample properties of the estimators of coefficients in the error components regression models. *Econometrica: journal of the Econometric Society*, 261–275.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B) 58*, 267–288.

Tweedie, M. C. (1957). Statistical properties of inverse gaussian distributions. i. *The Annals of Mathematical Statistics 28*(2), 362–377.

Umbach, S. L. (2020). Forecasting with supervised factor models. *Empirical Economics 58*(1), 169–190.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis 25*(1), 57–76.