Degree Project in The Built Environment
Second Cycle 30.0 credits

Urban Shrinkage Effects in Swedish Municipalities

A Random Forest Approach

HAUKUR ÖRN VALTÝSSON AND PONTUS FLYGRING

Title	Urban Shrinkage Effects in Swedish Municipalities: A Random Forest Approach
Authors	Haukur Örn Valtýsson and Pontus Flygring
Department	Real Estate and Construction Management
TRITA Number	TRITA-ABE-MBT-25247
Supervisor	Daria Kuznetsova
Key Words	Urban Shrinkage, Machine Learning, Random Forest, Housing Price, Population Change, Economic Restructuring

Abstract

Urban shrinkage has become a prominent global phenomenon, affecting cities across Europe, North America, and Asia due to long-term demographic decline, economic restructuring, and spatial realignment. These shifts increasingly challenge local governments as shrinking populations and underutilised housing stock strain planning, infrastructure, and service provision. Within this global context, Sweden illustrates these dynamics, with numerous municipalities experiencing population decline in recent years, exposing regional imbalances and housing market implications.

This study applies a data-driven approach to predict changes in population and housing prices across all 290 Swedish municipalities. Using publicly available data from Statistics Sweden (SCB), Random Forest regression models were developed for the periods 2016–2020 and 2019–2023. The models assess predictive accuracy and identify which demographic, economic, and housing variables most influence municipal-level change, based on Model Performance, Feature Importance analysis and Partial Dependence analysis.

Model results indicate that predictive accuracy is highest in larger, urban municipalities, whereas performance is more variable in smaller or rural areas, especially when modeling housing prices. Despite these differences, the models identify important structural relationships and key factors driving urban shrinkage. Rather than establishing causality, the models serve as anticipatory tools to inform evidence-based decision-making. The findings offer valuable insights for planners, policymakers, and real estate stakeholders, highlighting the practical and analytical potential of machine learning in understanding demographic shifts and market dynamics.

Titel	Urban krympning i svenska kommuner: En analys med Random Forest-metod
Författare	Haukur Örn Valtýsson and Pontus Flygring
Institution	Fastigheter och Byggande
TRITA Nummer	TRITA-ABE-MBT-25247
Handledare	Daria Kuznetsova
Nyckelord	Krympande kommuner, Maskininlärning, Random Forest, Bostadspris, Befolkning, Demografi, Ekonomisk omstrukturering

Sammanfattning

Befolkningsminskning i städer och kommuner har blivit ett allt mer påtagligt globalt fenomen som påverkar städer i Europa, Nordamerika och Asien till följd av bestående demografisk nedgång, omstrukturering av ekonomiska aktiviteter och förändrade stadsstrukturer. Dessa förändringar innebär växande utmaningar för lokala myndigheter, där befolkningsminskning och ett underutnyttjat bostadsbestånd försvårar planering, infrastruktur- och serviceförsörjning. I det globala sammanhanget är Sverige ett tydligt exempel, med flertalet kommuner som under de senaste åren upplevt kontinuerlig befolkningsminskning och underliggande regionala skillnader.

Denna studie använder en datadriven metod för att förutsäga förändringar i befolkning och bostadspriser i samtliga av Sveriges 290 kommuner. Med hjälp av offentligt tillgänglig statistik från Statistiska centralbyrån (SCB), utvecklades Random Forest-regressionsmodeller för perioderna 2016–2020 respektive 2019–2023. Modellprestanda utvärderas och de variabler (demografiska, ekonomiska och bostadsrelaterade) med störst inverkan på kommunala förändringar identifieras, baserat på varje variabels relativa betydelse i förhållande till övriga variabler i datasetet.

Modellerna presterade bäst i kommuner med större stadsbildning i analysen av befolkningsförändring, medan modellen för förändring av bostadspriser gav bättre resultat när de mindre kommunerna exkluderades. Trots vissa begränsningar påvisar modellerna viktiga strukturella samband och lyfter fram centrala indikatorer för befolkningsminskning. Modellerna är inte tänkta att förklara orsakssamband, utan används snarare som verktyg för att identifiera mönster och stödja databaserade beslut. Resultaten erbjuder insikter för planerare, beslutsfattare och aktörer inom fastighetssektorn och tydliggör det praktiska och analytiska värdet av maskininlärning i hanteringen av demografiska och marknadsdrivna förändringar.

Acknowledgements

This master thesis was completed during the spring of 2025 as part of the finalisation of the Real Estate and Construction Management program at KTH Royal Institute of Technology.

We would like to express our gratitude to our supervisor, Daria Kuznetsova, for their guidance, constructive feedback, and support throughout the course of our thesis work. A special thank you to Agnieszka Zalejska Jonsson for valuable suggestions and critical input that helped improve the quality of this thesis.

Finaly, we would also like to thank our peers who participated in the thesis seminars for their valuable feedback, insightful questions, and the inspiration they provided throughout the writing process.

Haukur Örn Valtýsson & Pontus Flygring
Stockholm, June 2, 2025

Förord

Detta examensarbete genomfördes under våren 2025 som en del av slutförandet av masterprogrammet i Fastigheter och Byggande vid Kungliga Tekniska högskolan (KTH).

Vi vill rikta ett stort tack till vår handledare, Daria Kuznetsova, för hennes vägledning, konstruktiva feedback och stöd under arbetets gång. Ett särskilt tack riktas även till Agnieszka Zalejska Jonsson för värdefulla synpunkter och kritisk återkoppling som bidrog till att höja kvaliteten på detta arbete.

Slutligen vill vi tacka våra kurskamrater som deltog i seminarierna för deras värdefulla kommentarer, insiktsfulla frågor och den inspiration de bidrog med under skrivprocessen.

Haukur Örn Valtýsson & Pontus Flygring
Stockholm, 2 Juni, 2025

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations and Key Concepts

RF – *Random Forest*: An ensemble machine learning method that builds multiple decision trees and aggregates their predictions to improve accuracy and reduce overfitting.

ML – *Machine Learning*: A field of artificial intelligence focused on algorithms that learn from data to make predictions or decisions.

Urban Shrinkage: A sustained decline in population, housing demand, and/or economic activity within a city or municipality.

PCR – *Population Change Rate*: A target variable measuring the percentage change in population over a defined period.

HPC – *Housing Price Change*: A target variable measuring the percentage change in average housing prices over a defined period.

DTT – *Demographic Transition Theory*: A theoretical model describing population change through stages of development (e.g., declining birth and death rates).

CCT – *Cumulative Causation Theory*: A theory describing how socio-economic decline reinforces itself through feedback mechanisms.

ULCT – *Urban Life Cycle Theory*: A model suggesting cities pass through phases of growth, maturity, and decline influenced by structural and spatial factors.

PDP – *Partial Dependence Plot*: A visualization tool for interpreting machine learning models by showing the marginal effect of a feature on the predicted outcome.

OOB – *Out-of-Bag*: A validation technique in ensemble learning where each tree is tested against data not used in its training sample.

Feature Importance: A ranking of variables based on how much they contribute to a model's predictions.

Hyperparameter Tuning: The process of optimizing model settings (e.g., number of trees, max depth) to improve performance.

Cross-sectional Data: Data collected at one point in time across multiple units (e.g., municipalities).

Threshold Effect: A point at which a small change in a variable result in a disproportionately large change in the outcome.

Nonlinearity: A relationship in which changes in one variable do not produce consistent, proportional changes in another.

Overfitting: A modelling issue where the model fits the training data well but performs poorly on unseen data due to excessive complexity.

1 Introduction

1.1 Background

Urban shrinkage has become a globally recognised pattern of spatial change for many large cities. Documented throughout European and North American industrial cities, it has since been observed in cities of varying size and character in Japan, Australia, Russia (Haase et al., 2014; Wiechmann and Pallagst, 2012), and the Nordic region (Grundel and Magnusson, 2023). According to Haase et al. (2014) urban shrinkage is a long-term, place-specific phenomenon driven by overlapping demographic (ageing and net out-migration), economic (deindustrialization and falling labour demand), and spatial (suburbanization and the relative decline of central urban areas) changes.

The structural changes are reflected in population loss, vacant spaces, and underutilised infrastructure, serving simultaneously as cause, context, and consequence within cycles of decline (Haase et al., 2014). These changes undermine municipal revenues (Eimermann et al., 2022) strain the provision of public services (Grundel and Magnusson, 2023), and destabilise housing markets. Turok and Mykhnenko (2007) further show that shrinkage follows multiple trajectories, shaped by local shocks, governance responses, and path-dependent development.

The phenomenon has been analysed across a wide range of disciplines. Demographic studies have examined trends such as population decline, ageing, and migration (e.g. Eimermann et al., 2022; Turok and Mykhnenko, 2007), while economic research has focused on labour-market contraction, structural unemployment, and fiscal stress (e.g. Bloom et al., 2010; He et al., 2017; Martinez-Fernandez et al., 2012). In parallel, real estate studies have documented oversupply, declining property values, and rising vacancy rates in shrinking cities (Eimermann et al., 2022; Haase et al., 2016).

In Sweden, recent demographic trends mirror broader EU patterns of stagnation and population ageing (Aurambout et al., 2021). During 2023 the national population of Sweden rose by just 30,200 people, the smallest absolute increase since 2001 (approximately 0.3 percent), driven by low birth rates, reduced net immigration and rising emigration (SCB, 2024a). Statistics Sweden (SCB), projected the total population to climb from 10,5 million at the end of 2023 to 11,8 million by 2070, only a 12 percent gain. In the same projection, the share of working-age residents (25-69 years) fell from 56 percent to 53 percent and the number of people aged 85 and above nearly tripled from around 280,000 to 750,000 (SCB, 2024a).

Yet, these national aggregates conceal significant spatial heterogeneity and sharp local imbalances across Sweden's municipalities. According to Boverket (2024), approximately half of Sweden's 290 municipalities now report an overall housing shortage, a marked decrease from 80 percent in 2023 to 51 percent in early 2024. These trends have created a growing mismatch between housing supply and demographic needs, particularly in areas with declining youth populations and ageing residents who are more likely to remain in their homes as they grow older (Boverket, 2024; Eimermann et al., 2022).

This mismatch reduces confidence in the housing market as public planners and private investors must navigate both financial uncertainty and potential policy challenges (Grundel and Magnusson, 2023; Haase et al., 2014). Smaller municipalities in northern and inland regions are particularly exposed, facing compounding pressures from structural changes accompanied by labour-market contraction (Eimermann et al., 2022). Grundel and Magnusson (2023) argue that conventional infrastructure-planning frameworks often overlook these uneven dynamics, applying one-size-fits-all solutions unsuited to local shrinking contexts, risking over-investment in declining regions and underinvestment in areas of emerging need. This highlights the need for locally sensitive, empirically grounded tools to inform more adaptive and locally appropriate planning responses.

1.2 Research Problematisation & Gap

Urban shrinkage reshapes how demographic, economic, and spatial forces interact at the municipal-level (Haase et al., 2014). However, much of the existing literature take a largely descriptive or retrospective approach, often focusing on case studies of population decline or post-hoc governance responses (Grundel and Magnusson, 2023; Khavarian-Garmsir, 2023; Wiechmann and Pallagst, 2012). Though valuable in tracing past trends, such studies lack the predictive methods needed for informed, proactive decisions, few studies model how variables from multiple domains interact to reinforce shrinkage. Recent literature highlights the need for predictive and integrated analytical frameworks capable of capturing complex non-linear relationships that linear analysis often overlook (Peng et al., 2023).

Although predictive machine-learning methods such as Random Forests have been applied to urban shrinkage forecasting in Japan and other international contexts (Peng et al., 2023; Wang et al., 2024), similar approaches remain absent from the Swedish literature. Existing studies in Sweden do not combine demographic, economic, and housing indicators in predictive models of population or housing-market change. This likely reflects a continued reliance on single-domain frameworks and a limited usage of predictive modelling approaches. This leaves nationally available data, such as that provided by SCB and Boverket, largely underused in integrated, forward-looking analyses of municipal dynamics.

To address this gap, the study develops two cross-sectional Random Forest models covering all 290 Swedish municipalities across two periods, 2016-2020 and 2019-2023, where both models incorporate variables from the three domains. One model focuses on Population Change Rate (PCR) while the other targets Housing Price Change (HPC). The scope and delimitations of the study is presented in section 1.4, and a detailed explanation of the two models and a complete list of variables is provided in Chapter 4.

1.3 Research Objective and Questions

This study develops and applies Random Forest (RF) regression models to predict two outcome variables related to urban shrinkage at the municipal-level in Sweden: population change rate (PCR) and housing-price change (HPC). The aim is to (1) evaluate the predictive performance of these models and (2) uncover nonlinear relationships between predictors and outcomes, including the structural and temporal complexities of urban shrinkage.

This is addressed through the following research questions:

1. *How accurately can changes in population and house prices in Swedish municipalities be predicted using cross-sectional Random Forest regression using demographic, economic, and housing indicators, and how does model performance compare?*

2. *What do the models reveal about the nonlinear relationships between predictor variables and outcome variables and the structural and temporal complexities of urban shrinkage?*

These questions are addressed through two Random Forest models applied to the cross-sectional periods 2016–2020 and 2019–2023. Including both the PCR and HPC models in the study enables a comparative analysis that allows for analysing the feedback loops between demographic and housing-market dynamics at the municipal-level, interpreted through the theoretical framework outlined in Section 2.7.

1.4 Delimitations and Scope

This study focuses on modelling municipal shrinkage in Sweden using a cross-sectional, data-driven approach. It is limited to structural indicators, such as demographic composition, economic conditions (e.g., unemployment rate), and housing market characteristics (e.g., dwelling stock change), that are publicly available at the municipal-level. The data for the model variables are exclusively drawn from SCB and are on a annual municipal-level unless other is stated. The analysis uses data from two defined periods: 2016–2020 and 2019–2023. These periods were selected to capture structural shifts in pre- and post-COVID housing market adjustment, offering contrasting macroeconomic and demographic contexts for analysis. All variables are aggregated at the municipality scale; intra-municipal variation is not considered.

While the study models both population change and housing price development, it does so by a cross-sectional design that captures structural differences between municipalities at two distinct time points. This approach prioritises comparability over time series modelling, meaning that within-municipality changes, short-term fluctuations, and policy effects are not captured. Although longitudinal or panel methods could provide deeper insight into temporal dynamics and delayed outcomes, such approaches would require multi-decade data coverage and is beyond the scope of this study. Rather than producing forward-looking forecasts, the models evaluate structural associations and predictive patterns within defined historical windows, offering a basis for comparative analysis while acknowledging the limits of temporal generalisation.

Both the PCR model and the HPC model, use the same time periods for consistency and comparability. This shared framework allows for direct assessment of how structural indicators relate to different forms of urban shrinkage under similar external conditions. While population change is often recorded more immediately through migration and demographic statistics, housing market indicators such as prices and vacancies typically respond with a lag due to market frictions, transaction delays, and investment cycles. Using the same periods for both models enables the analysis to capture these cumulative effects, while maintaining a coherent basis for evaluating how the same indicator domains behave across outcome types. This consistency strengthens internal validity and allows for more structured interpretation of variable importance across domains. All variables used in the analysis are numeric or were transformed into compatible formats suitable for Random Forest regression. Variable preprocessing steps, including the treatment of missing values, encoding of categorical data, and scaling where relevant, are detailed in Chapter 4.

The study adopts a non-parametric predictive framework that does not rely on predefined functional relationships between variables. Random Forest regression was selected over traditional econometric approaches (e.g. OLS or fixed effects models) due to its robustness against overfitting and in handling high-dimensional data, tolerance for multicollinearity, and its ability to detect complex non-linear interactions without assuming variable distributions. Accordingly, the analysis does not attempt to establish causal relationships or evaluate the impact of specific policy interventions. Causal inference would require experimental variation or longitudinal panel structures that are beyond the scope of this design. Interpretations focus on predictive accuracy and the relative contribution of variables to model performance, as measured by feature importance scores.

The study also does not incorporate qualitative indicators such as governance strategies, institutional quality, or local planning decisions, which may also influence shrinkage outcomes. These factors are excluded due to limitations in availability, comparability, and the structural modelling scope of this thesis. This thesis also does not address environmental or climate-related aspects of urban shrinkage, such as land-use shifts due to flood risk or sustainability transitions in shrinking regions. While these themes are increasingly discussed in the literature, they fall outside the structural, quantitative scope of this study.

Although the national dataset includes all 290 municipalities, selected analyses, particularly those involving housing price modelling, also test subsamples based on population size thresholds. Stratification by municipal size allows the study to examine whether model performance differs systematically between small, medium, and large municipalities, an important consideration in the Swedish context, where shrinkage dynamics vary regionally. However, the findings from these subsamples are not intended to be extrapolated beyond the studied groups.

2 Literature Review

2.1 Urban Shrinkage: Concepts and Definitions

An early introduction of the Urban Shrinkage was made by Häußermann and Siebel in 1988 (Ma et al., 2024) to describe the sustained population loss and economic decline observed in industrial cities. Since then, the concept has gained growing attention, prompting researchers to explore how shrinkage is defined, measured, and understood across various spatial and temporal contexts (Beauregard, 2009; Haase et al., 2014; Martinez-Fernandez et al., 2012; Reckien and Martinez-Fernandez, 2011). In urban shrinkage research, population decline is the most commonly cited indicator, owing to its accessibility and comparability (Doeringer et al., 2020), a growing consensus emphasizes that it should not serve as a stand-alone definition. Instead, urban shrinkage should be seen as a complex phenomenon shaped by demographic, economic, and spatial dynamics (Haase et al., 2014; Martinez-Fernandez et al., 2012).

Schilling and Logan (2008) define a shrinking city as an old industrial urban area that has lost 25 percent of its population over a 40-year period. The Shrinking Cities International Research Network (SCIRN) offers a broader definition, identifying shrinking cities as densely populated areas with more than 10,000 inhabitants experiencing economic transformation, structural crisis, and population loss over a period spanning more than two years (Wiechmann, 2008). These definitions provide useful starting points but also raise important methodological and conceptual concerns. For example, the 25 percent threshold proposed by Schilling and Logan (2008) may be too rigid or context-specific to apply across different urban contexts. Cities undergoing slower, less dramatic forms of decline or those affected by subtle demographic shifts may be overlooked.

To address these limitations, literature advocates for a multidimensional understanding of urban shrinkage, one that goes beyond population metrics (Doeringer et al., 2020; Liu et al., 2020). For instance, the definition provided by SCIRN emphasise density and size but may exclude urban areas organised around multiple interconnected urban centres, or smaller towns facing similar dynamics (Wiechmann, 2008). Doeringer et al. (2020) argue that shrinkage manifests in varied forms depending on local context and evolves differently across time. In some cities, it occurs abruptly, while in others, it unfolds gradually over decades, particularly in regions undergoing demographic transition. This highlights the need for approaches to urban shrinkage that integrate measurable indicators with theoretical perspectives capable of addressing both structural trends and local conditions. Rather than representing a fixed endpoint, urban shrinkage is increasingly understood as a dynamic process shaped by feedback mechanisms and structural change (Doeringer et al., 2020; Haase et al., 2014). Building on this perspective, Haase et al. (2014) propose an integrative model in which shrinkage unfolds through interconnected phases. Including root causes (e.g., deindustrialization, demographic shifts), resulting impacts (e.g., vacant housing, fiscal decline), policy responses, and feedback loops.

Once considered a locally specific phenomenon linked to post-industrial cities in Europe and North America, urban shrinkage now increasingly affects urban areas across Asia, Africa, and Latin America (Aurambout et al., 2021; Martinez-Fernandez et al., 2012). Moreover, population decline is no longer confined to economically struggling cities. Projections indicate that even cities currently experiencing growth may face future shrinkage due to aging populations and declining birth rates (Aurambout et al., 2021). Despite the growing research interest, no unified theoretical framework fully explains urban shrinkage across different contexts, as the empirical evidence remains fragmented (Kato, 2024; Khavarian-Garmsir, 2023).

2.2 Economic Restructuring and Urban Shrinkage

Economic change has disproportionately favoured large metropolitan areas, concentrating investment and job opportunities in a few urban centres while many smaller cities and rural municipalities have experienced industrial decline and disinvestment (Martinez-Fernandez et al., 2012; Wiechmann and Pallagst, 2012). These spatial disparities underscore the uneven foundations of urban shrinkage, linking long-term demographic decline to broader structural transformations in the economic landscape. As Haase et al. (2014) emphasise, economic decline often triggers out-migration and population loss, while demographic contraction weakens the economic base required to sustain services and attract investment. This feedback mechanism reinforces long-term shrinkage, making recovery increasingly difficult once decline is underway. Empirical studies from Europe and the United States show that cities affected by early industrial decline frequently face persistent population loss and reduced investment capacity (Wiechmann and Pallagst, 2012), illustrating the lasting structural consequences of economic restructuring.

As Martinez-Fernandez et al. (2012) note, globalisation is also condensing economic activities, intellectual assets, and infrastructural investments in a number of "global cities," building on the work of Sassen (2001). This restructuring has left many smaller and mid-sized urban areas facing capital outflows, job losses, and economic decline, contributing to widening regional divides between expanding metropolitan centres and contracting peripheral areas (Martinez-Fernandez et al., 2012).

Sweden's transition toward a service- and knowledge-based economy particularly in sectors such as information technology, finance, and professional services, has amplified the demographic divides. High-skill employment and investment have become increasingly concentrated in metropolitan regions while smaller municipalities continue to face economic disadvantages, labour market stagnation, and youth outmigration (Hedlund and Lundholm, 2015; Tillväxtverket, 2024).

Beyond labour markets, the economic consequences of shrinkage extend to local fiscal capacity and infrastructure sustainability. As Saraiva et al. (2017) demonstrate, population loss linked to economic restructuring can depress municipal tax bases, raise per capita service burdens, and jeopardize the viability of public infrastructure. Similar dynamics have been observed in the Swedish context (Eimermann et al., 2022; Grundel and Magnusson, 2023; Lundmark et al., 2022). Acknowledging these interdependencies highlights the limitations of relying solely on demographic models to understand urban shrinkage. Labour market shifts, investment flows, and structural economic vulnerability should therefore be considered within a comprehensive framework (Grundel and Magnusson, 2023; Haase et al., 2014).

2.3 Global and European Demographic Trends

Urbanization, once synonymous with continuous growth, has become increasingly uneven and uncertain. According to the United Nations, Department of Economic and Social Affairs (UN DESA), 63 countries, including many in Europe, have already surpassed their population peaks and are projected to shrink by an average of 14 percent over the next 30 years. Birth rates have fallen below the replacement level of 2.1 children per woman in more than half of all countries, with nearly one-fifth experiencing "ultra-low" birth rates below 1.4. Simultaneously, the global population is ageing, with the number of people aged 65 and older expected to exceed those under 18 by the late 2070s (UN DESA, 2024). These demographic transitions, long-term shifts from high to low birth and death rates, are driving aging populations and population decline, with implications for labour supply, fiscal sustainability, and spatial planning (Bloom et al., 2010).

Between 2001 and 2011, more than a quarter of the EU's Functional Urban Areas (FUAs)—urban regions defined by commuting patterns and socio-economic integration—experienced population decline. A trend that persisted through 2018. By 2050, an estimated 45 percent of FUAs could be experiencing population decline, affecting nearly a quarter of the region's urban residents. Some cities particularly in Germany, Latvia, Lithuania, and Bulgaria are projected to lose more than 20 percent of their 2020 population (Aurambout et al., 2021).

However, the patterns of shrinkage within Europe are heterogeneous. Aurambout et al. (2021) identify three primary trajectories: cities that continuously shrink, cities that rebound after periods of shrinkage, and cities that shift from growth to decline following economic shocks such as the 2008 financial crisis. This distinction parallels findings by Martinez-Fernandez et al. (2012) who emphasize that shrinkage may be either temporary or structural, depending on broader demographic, economic, and institutional trajectories. Turok and Mykhnenko (2007) similarly categorize European cities into distinct demographic trajectories, ranging from continuous growth to various forms of decline and resurgence, reinforcing the view that urban shrinkage is rarely linear and often cyclical. These findings highlight how institutional arrangements and regional development histories shape national-level differences in urban shrinkage trajectories. Importantly, their analysis shows that sustained resurgence is relatively rare, with a far greater number of cities experiencing long-term or recent population downturns (Turok and Mykhnenko, 2007).

Suburbanisation introduces an additional layer of complexity, as population decline in central urban areas can coexist with expansion in suburban and peri-urban zones (areas on the outskirts of cities that often blend urban and rural characteristics), resulting in fragmented and spatially dispersed patterns of shrinkage within single metropolitan regions (Haase et al., 2014; Wiechmann and Pallagst, 2012). These spatial complexities are mirrored by international variation in shrinkage drivers. Comparative perspectives reveal that while shrinkage in Europe is often driven by economic decline and structural adjustment, in Japan it is primarily attributed to demographic ageing and ultra-low birth rates (Doeringer et al., 2020; Wiechmann and Pallagst, 2012).

Taken together, the demographic, economic, and spatial transformations mentioned in previous sections create a fragmented and uneven foundation for understanding urban shrinkage. Taken together, the demographic, economic, and spatial transformations mentioned in previous sections create a fragmented and uneven foundation for understanding urban shrinkage. These dynamics point to the value of anticipatory and locally adapted governance strategies, particularly in countries like Sweden, where municipal growth and decline coexists (Eimermann et al., 2022).

2.4 Urban Shrinkage in the Swedish Context

Sweden illustrates how urban shrinkage unfolds unevenly within a high-income, decentralised welfare state. National trends such as falling fertility, population aging, and growing reliance on immigration are not evenly distributed across the country. Rural and northern municipalities are disproportionately affected by population loss and out-migration, compounded by structural economic and spatial disadvantages

(Boverket, 2024; Eimermann et al., 2022). These internal disparities are often obscured by national-level statistics.

According to SCB, the country's birth rate fell to 1.45 children per woman in 2023, the lowest level recorded since national statistics began in 1749. Natural population growth is projected to turn negative by the early 2030s, and without sustained high levels of immigration, overall growth is expected to decline (SCB, 2024b). One projected consequence is a decline of nearly 100,000 children aged 6 to 15 by 2070, posing substantial planning challenges for shrinking municipalities, particularly in sustaining school infrastructure and service provision. By 2070, nearly one-quarter of Sweden's population is projected to be foreign-born, underscoring the growing role of migration in offsetting long-term demographic decline. Median ages in shrinking municipalities now surpass 45 years, whereas metropolitan cores such as Stockholm and Gothenburg report median ages closer to 38 years (SCB, 2024a). Internal migration continues to widen regional disparities, as young adults increasingly move from peripheral municipalities to urban labour markets (Carson et al., 2016; Eimermann et al., 2022; Tillväxtverket, 2024).

Åsele Municipality in Västerbotten County provides a compelling example of the spatial and demographic complexities. While the municipality has experienced long-term population decline, Eimermann et al. (2022) show that these trends mask significant sub-municipal variation. For instance, the small village of Gafsele has attracted lifestyle migrants from outside of Sweden, drawn by outdoor amenities and niche activities. These localised growth pockets coexist with broader municipal demographic decline, revealing tensions between declining public services and unanticipated demands from new residents. This unplanned growth also reflects institutional blind spots in planning frameworks, which often fail to anticipate non-traditional migration patterns in shrinking regions (Grundel and Magnusson, 2023).

Governance responses to urban shrinkage in Sweden reveal a disconnect between formal municipal autonomy and practical financial dependency. Although local governments are politically independent, they rely heavily on intergovernmental transfers and national welfare systems to maintain services (Grundel and Magnusson, 2023). This dependency becomes increasingly problematic as shrinking tax bases and rising per capita costs force municipalities to cut services, defer infrastructure maintenance, or restructure administration. These tensions are further exacerbated by fragmented regional planning and national growth discourses that often overlook intra-regional disparities Lundmark et al. (2022). Many municipalities continue to adopt growth-oriented strategies, not due to realistic growth prospects, but because institutional lock-ins constrain the ability to abandon inherited planning models. While recent national strategies, such as *Ny strategi för levande och trygga städer*

(Regeringen, 2025), acknowledge the implications of demographic decline, implementation remains uneven. Regional policy documents likewise highlight growing vulnerability among northern and inland municipalities, alongside mounting pressures in expanding urban regions (Boverket, 2024; Tillväxtverket, 2024). These governance challenges also reveal key institutional contrasts with Sweden's Nordic neighbours. Unlike Denmark and Norway, where planning mandates are tightly coordinated across governance levels (Nordregio, n.d.), Sweden's fragmented vertical integration limits municipalities' capacity to anticipate and manage demographic shifts (Grundel and Magnusson, 2023; Nordregio, n.d.).

Taken together, Sweden's demographic patterns, institutional arrangements, and regional disparities make it a particularly instructive case for modelling urban shrinkage at the municipal-level. The country offers both extensive data and pronounced intra-national variation, enabling analysis of how structural dynamics interact across different local contexts. This variation justifies a cross-sectional modelling approach capable of capturing municipality-level divergence within a single national setting. The Swedish case highlights how municipality-level variation can offer critical insights into the structural foundations of urban shrinkage, with relevance for both research and policy in high-income countries.

2.5 Housing Market Dynamics in Shrinking Areas

Housing markets both reflect and amplify urban shrinkage. Declining populations, demographic aging, and economic restructuring alter patterns of housing demand, supply, and investment across shrinking and growing municipalities. In shrinking areas, reduced household formation and the out-migration of working-age residents often produce a growing mismatch between the existing housing stock and current demographic needs (Saraiva et al., 2017). This dynamic is reinforced by long-term demographic changes, which have been shown to depress housing demand and price growth even in relatively stable economic conditions (Malmberg, 2010).

National housing market data reinforce these trends within the Swedish context. According to Boverket's (2024) Housing Market Survey, 119 of 288 municipalities now classify their markets as balanced—an increase of 22 since 2023 and 53 since 2022. Of these, the majority are smaller municipalities with fewer than 25,000 inhabitants, although shortages remain widespread, particularly for financially vulnerable groups such as pensioners and young adults. Vacancy rates now exceed 10 percent in many shrinking municipalities, compared to under 2 percent in growing metropolitan areas (Boverket, 2024). These imbalances risk evolving into long-term structural vacancies, complicating land use planning, infrastructure maintenance, and fiscal sustainability (Eimermann et al., 2022). At the same time, new construction remains limited in shrinking areas, as subdued demand and financial risk deter both public and private sector investment. Empirical studies confirm that demographic composition, particularly the shrinking share of younger adults, has a strong and predictive influence on residential construction levels in Sweden (Lindh and Malmberg, 2008). Although large-scale industrial investments in northern Sweden have generated localised job growth, they have not reversed broader demographic or housing market decline, highlighting the fragmented and selective nature of contemporary re-industrialisation (Lundmark et al., 2022). These patterns underscore the widening structural divide between housing markets in growing and shrinking municipalities (Eimermann et al., 2022).

In shrinking municipalities, larger detached houses in peripheral areas, once suited to younger families, are increasingly misaligned with the needs of aging, single-person households (Boverket, 2024). As demand shifts toward smaller, more accessible housing in central locations, limited financial incentives and weak market signals make it difficult for municipalities to retrofit or adapt their existing stock accordingly (Eimermann et al., 2022; Grundel and Magnusson, 2023). While some European countries, notably Germany, have implemented coordinated "shrink smart" strategies involving selective demolition, repurposing, or targeted redevelopment (Martinez-Fernandez et al., 2012), such approaches remain largely absent in the Swedish context. As Eimermann et al. (2022) illustrate through the case of Åsele Municipality, these dilemmas are not only technical or fiscal but also politically and socially charged. Municipalities are often left to manage surplus housing and associated infrastructure burdens independently, placing additional pressure on already stretched local budgets and planning capacities (Grundel and Magnusson, 2023).

Recognizing the structural role of housing markets is essential for understanding shrinkage, especially in later stages of the urban life cycle when declining demand and aging stock deepen spatial and fiscal vulnerabilities (Haase et al., 2014).

2.6 Machine Learning Applications in Urban and Regional Studies

Classical econometric models, such as linear regression using Ordinary Least Squares (OLS), rely on strict assumptions: linear relationships, independence of variables, and constant variance of residuals (homoscedasticity). These conditions are rarely met in urban systems, which are characterised by feedback loops, heterogeneity, and spatial dependence (Breiman, 2001; Shmueli, 2010). As a result, Machine Learning (ML) methods have gained traction in urban and regional studies, offering more flexible, robust, and accurate tools for capturing complex relationships in large, noisy, and high-dimensional datasets (Hastie et al., 2009).

Among ML techniques, ensemble models, which combine the predictions of multiple weaker learners to improve overall accuracy, such as Random Forests (RF) have been effective in addressing the challenges posed by urban and regional data (Breiman, 2001; Peng et al., 2023). Introduced by Breiman (2001), RF aggregate the predictions of multiple decision trees, constructed by sampling the training data and selecting random subsets of predictors, to achieve high predictive performance while minimising overfitting. RF offer a combination of strengths, the possibility to handle multicollinearity without pre-processing, detect complex interactions automatically, and generate transparent measures of variable importance (Hastie et al., 2009). RF balance the bias-variance trade-off by reducing overfitting while preserving complexity in high-dimensional, heterogeneous datasets. (Hastie et al., 2009).

Peng et al. (2023) applied RF to model urban shrinkage in Japan, demonstrating the method's ability to capture nonlinear influences of demographic, economic, and spatial factors on municipal population dynamics. Similarly, Wang et al. (2024) utilized RF to forecast material stock changes in Japanese cities under demographic decline, showcasing ML's effectiveness in modelling urban sustainability challenges. In the real estate domain, Ma et al. (2020) employed RF and Gradient Boosting methods to model land values based on spatial and economic predictors, highlighting the advantages of ML in capturing complex urban economic phenomena. This growing relevance is further reflected in recent systematic reviews (Wang and Biljecki, 2022), which highlight the expanding use of ML methods in urban studies.

Despite growing international adoption of RF models in urban shrinkage research, their application within the Swedish context remains limited. This gap underscores the need for modelling approaches tailored to Sweden's specific institutional, spatial, economic, and demographic conditions.

3 Theoretical Framework

While this thesis primarily employs a predictive modelling approach, grounding the analysis in theoretical perspectives is essential for shaping the analytical approach, guiding variable selection, and situating findings within a wider academic context. As urban shrinkage cannot be understood through single-factor explanations but requires integrating multiple overlapping processes (Haase et al., 2014; Martinez-Fernandez et al., 2012; Reckien and Martinez-Fernandez, 2011). This section presents three complementary theoretical frameworks that provide a foundation for understanding the drivers of shrinkage and framing the subsequent empirical work: Demographic Transition Theory, Cumulative Causation Theory, and Urban Life Cycle Theory. Their relevance and limitations are critically discussed, with particular attention to how they shape the overall conceptual framing and inform the interpretation of patterns observed through the machine learning approach.

3.1.1 Demographic Transition Theory

Demographic Transition Theory (DTT) explains how societies progress through stages of population change, beginning with high birth and death rates and eventually reaching low birth and mortality, resulting in population stagnation and potential decline (Kirk, 1996; Lee, 2003). In its final stage, low birth rates paired with increased life expectancy produce aging populations and negative natural population growth, phenomena closely linked to urban shrinkage, particularly in advanced economies.

Sweden exemplifies this demographic stage, with birth rates below replacement level and a rapidly aging population (SCB, 2024a). Variables such as age dependency ratios, birth rates, and migration patterns serve as key structural inputs for evaluating shrinkage risk (Eimermann et al., 2022; Haase et al., 2014). However, DTT primarily addresses macro-level transitions and tends to generalize across spatial scales, which can obscure the diversity of municipal trajectories. It does not adequately capture localized shocks, migration flows, or institutional responses that shape urban shrinkage at the municipal-level. As such, while DTT offers a valuable structural backdrop, it benefits from being complemented by frameworks that account for spatial and institutional variation.

In this study, DTT is used not as a causal framework but as a framing tool to interpret demographic variables as part of broader structural change. It supports the view that population decline is a systemic, path-dependent process rather than a series of isolated events, thereby providing interpretive context for the demographic indicators incorporated into the predictive modelling framework.

3.1.2 Cumulative Causation Theory

Cumulative Causation Theory (CCT), developed by Gunnar Myrdal has been used to conceptualise urban shrinkage as a dynamic, self-reinforcing process driven by feedback loops that amplify initial demographic or economic shocks (Myrdal, 1960). For example, population loss reduces municipal tax revenues, leading to service cuts and infrastructure underuse, which further discourages investment and triggers additional outmigration (Haase et al., 2014).

CCT incorporates the notion of path dependence, where historical conditions and, in some contexts, institutional structures may lock municipalities into trajectories of decline or stagnation (Myrdal, 1960). An initial advantage, or disadvantage, in one region or sector (for example, higher investment, productivity gains, or infrastructure quality) generates further benefits (such as greater labour-force skills, capital inflows, and market access), which in turn attract even more resources and activity. In the Swedish context, institutional fragmentation and limited vertical integration between national strategies and local implementation (Grundel and Magnusson, 2023; Nordregio, n.d.) can reinforce the effects of cumulative causation. Areas that start off disadvantaged can become "locked in" to downward trajectories as out-migration, underinvestment, and institutional neglect compound one another (Myrdal, 1960). In spatial terms, it explains persistent differences between growing metropolitan cores and peripheral or shrinking areas, highlighting the role of historical path dependence, spillovers, and institutional structures in shaping long-run development outcomes.

While CCT provides a useful lens for understanding the persistence and intensification of decline, it may not fully account for the potential influence of policy interventions or local adaptive strategies that could enable recovery or transformation. Random Forest regression approximate feedback effects through partial dependence and importance scores (Breiman, 2001), but cannot fully capture causal loops. CCT thus informs the interpretation of variable interactions and model outputs, helping to identify patterns most closely linked to population change and decline.

3.1.3 Urban Life Cycle Theory

Urban Life Cycle Theory (ULCT) suggests that cities often follow phases of growth, maturity, and decline, characterised by changing population dynamics, economic activity, and spatial organisation (Martinez-Fernandez et al., 2012; Wiechmann and Pallagst, 2012). During growth, demand for housing and services expands, in maturity, markets stabilize, and in decline, population loss and economic contraction lead to housing surpluses, rising vacancies, and price depreciation (Haase et al., 2014).

ULCT provides a temporal-spatial lens to understand variation across municipalities in shrinkage intensity and housing market outcomes across the two periods. These dynamics underscore the role of housing market indicators as key proxies for identifying municipalities in later stages of the urban life cycle. As observed in Sweden's rural municipalities (Boverket, 2024), reflect how real estate feedback loops both signal and reinforce urban shrinkage.

While the life cycle framing serves as a helpful conceptual lens, it has been critiqued for reinforcing linear and deterministic assumptions that may not reflect the complexity of real-world shrinkage dynamics (Wiechmann and Pallagst, 2012). Urban trajectories are often nonlinear and spatially uneven, with some municipalities demonstrating resilience or transformation despite broader decline. Still, ULCT complements DTT and CCT by framing housing market signals as expressions of urban phase transitions.

3.1.4 Synthesis and Application

Together, Demographic Transition Theory, Cumulative Causation Theory, and Urban Life Cycle Theory offer a multidimensional conceptual framework for understanding urban shrinkage as the product of interconnected urban dynamics. DTT contextualises demographic pressures driving municipal change, CCT highlights feedback-driven persistence rooted in economic decline and reinforced by institutional and spatial context, and ULCT interprets housing indicators as markers of cities' positions within broader urban trajectories. This framework guides the selection of variables incorporated into the Random Forest models and supports interpretation model output. It also serves as a conceptual bridge between the literature and the empirical method presented in Chapter 4, helping to ensure that the predictive approach is informed by the underlying theoretical context.

While the core focus remains on structural indicators, it is also important to acknowledge that urban shrinkage is shaped by spatial externalities. Decline in one municipality may influence surrounding areas, reinforcing regional shrinkage clusters (Turok and Mykhnenko, 2007). Though not directly modelled here, this dynamic underscore the potential relevance of future spatial extensions.

Importantly, while these theories clarify underlying processes, the study remains predictive rather than explanatory. The machine learning approach captures nonlinear relationships consistent with these theoretical insights, enabling a data-driven assessment of structural patterns across Swedish municipalities.

4 Method

This chapter outlines the data-driven research strategy used to analyse municipal-level trends in population change and housing market development. The chapter begins with a description of the data foundation and variable logic, then details the machine learning setup, validation process, and interpretive techniques, before concluding with a critical reflection on data limitations and ethical considerations.

4.1 Data

Sweden's statistical transparency and decentralized governance create a robust data environment for analyzing urban and regional dynamics at the municipal-level. High-quality, granular, and regularly updated datasets are systematically collected and shared by both national agencies and private-sector platforms. A key example is SCB's Statistical Database, which provides comprehensive datasets across a wide array of domains. These data are standardized and updated regularly, ensuring both comparability and longitudinal depth (SCB, 2024c). The reliability and accessibility of Sweden's public data infrastructure have been internationally recognized, with the country ranked among global leaders in digital governance and open data provision (OECD, 2019).

Sweden's decentralized administrative structure and strong commitment to open-data principles have fostered a detailed, accessible public data ecosystem (OECD, 2023, 2019). This robust infrastructure provides a solid foundation for machine learning applications.

4.1.1 Data Sources

The data cover two structurally comparable four-year periods, 2016–2020 and 2019–2023, selected to support comparative analysis while maintaining a clear cross-sectional design. All data are secondary, sourced from SCB's official administrative and statistical records via the public Statistical Database platform. SCB data are compiled from official administrative registers and national surveys, ensuring a high level of consistency and comparability across municipalities. Definitions and measurement formats remained stable across the two periods. Variables are reported in a mix of absolute values, ratios, and standardized forms (e.g., per 1,000 residents), depending on the variable. Details regarding transformation and preprocessing are provided in Section 4.1.3.

4.1.2 Variable Selection

The selection of variables for this study was guided by both theoretical relevance and data availability. As outlined in Chapter 3, the theoretical framework identifies three interlinked domains—demographic, economic and housing—as central to understanding urban shrinkage. These domains inform the initial inclusion criteria for variables. Subsequently, an iterative, exploratory process refined the variable set. Each candidate variable was evaluated for municipal-level availability across both study periods and for empirical relevance during preliminary modeling. When multiple variables represented similar constructs, we retained those with the most complete data coverage and the greatest interpretability. This approach balances strong conceptual grounding with empirical tractability.

4.1.3 Variable Construction and Pre-processing

All data were harmonized at the municipal-level using standardised municipal codes, enabling accurate merging across variables and ensuring consistent alignment.

Table 1 lists all predictor variables used in the RF regression models along with their descriptions. To account for differences in municipality size, scale-sensitive variables, such as construction activity, were normalised per 1,000 residents at the start of each period. This adjustment ensures that comparisons across municipalities reflect underlying conditions rather than population magnitude. Housing supply was represented using the best available proxies: completed dwellings, started dwellings, and total dwelling stock. While these variables do not fully capture the complexity of local construction dynamics (e.g., planning timelines, investor behavior), they offer a practical approximation of supply-side conditions and retain coverage of the domain. Several additional variables were constructed to reflect structural signals relevant to urban shrinkage. A rent-to-price (RTP) ratio was derived as a proxy for rental affordability and hence gives context to rent levels compared to the local housing market.

Dwelling stock change was computed as the net delta in housing units over the period. To capture dynamic effects, momentum variables were created for house prices, disposable income, and dwelling stock, reflecting the three-year rate of change prior to each period. Expected living age, reported by SCB as four-year moving averages (e.g., 2010–2014), was aligned to the study periods by taking the median of values from the two overlapping brackets. Additionally, a proxy for industrial employment concentration was constructed by calculating the share of employed individuals within each municipality working in the manufacturing sector. This variable captures the relative size of the industrial labour market and serves as an indicator of structural economic composition. Given the well-documented link between deindustrialisation and urban shrinkage in the literature (Aurambout et al., 2021; Martinez-Fernandez et al., 2012), this metric provides a useful proxy for assessing exposure to long-term economic restructuring.

Variable selection was shaped through an iterative modelling process. Multiple candidate variables were initially tested and subsequently excluded based on limited predictive contribution or multicollinearity. While Random Forests are tolerant of multicollinearity, this filtering helped sharpen model interpretability and maintain conceptual balance across domains.

Table 1: Variable list with definitions

Variable	Description
AHP	Average Household Size (persons per household)
ALE	Average Life Expectancy (years)
FPR	Foreign-Born Population Rate (%)
MR	Male Share (%)
ODR	Old-Age Dependency Ratio (65+/15–64)
PD	Population Density (persons per km² of inhabitable land)
TP	Total Population (persons)
UPR	Underage Population Rate (under 19%)
RTP	Rent-to-Price Ratio (SEK/m² per house price, in 1000s)
GDP_pc	GDP per Capita (SEK)
ER	Employment Rate (% of total labour force)
PCR_lag5	Population Change Rate (Lag 5 Years) (%)
INC	Disposable Income per Capita (SEK)
MS	Manufacturing Employment Share (%)
INC_lag3	Income Growth (Lag 3 Years) (%)
HPC_lag3	House Price Change (Lag 3 Years) (%)
SD_1000	Started Dwellings per 1000
DSC_lag3	Dwelling Stock Change per 1000 Inhabitants (Lag 3 Years)

4.1.4 Missing Values

Variable datasets were nearly complete for all municipalities across both time periods. A small number of missing values occurred in SCB's rent-per-square-meter dataset (used to derive RTP): 28 cases in 2016, 6 in 2019, and one in 2020. These missing values were imputed using a Random Forest-based procedure that leverages rent-per-square-meter observations from other years to estimate the missing entries.

4.1.5 Outlier Treatment

No formal outlier detection or removal procedures were applied to any of the variables. Given the robustness of RF to extreme values (Breiman, 2001; Hastie et al., 2009), and the analytical importance of capturing the full range of municipal variation (including cases of unusually sharp population decline or growth) all municipalities and their associated values were retained in the modelling process.

4.2 Machine Learning Framework

This section outlines the implementation of Random Forest regression models used to predict municipal-level population change and housing price development. Two target variables were modeled separately: the percentage change in total population (Population Change Rate, PCR) and the percentage change in housing prices over each cross-sectional period.

For each of the two outcome variables, separate RF models were trained. This dual-model structure enables the model to examine how structural indicators predict different dimensions of urban shrinkage. Random Forests were selected over alternative modelling approaches—including Ordinary Least Squares regression, logistic regression, and Support Vector Machines—for its ability to handle non-linear patterns and maintain robustness to noise. Compared to boosting methods such as Gradient Boosting Machines and XGBoost, RF offers similarly strong predictive performance with fewer tuning requirements and greater stability in noisy settings (Hastie et al., 2009). In addition, RF provides built-in diagnostics for assessing generalisation error via out-of-bag (OOB) validation and feature importance (Breiman, 2001; Hastie et al., 2009).

Model performance during tuning was estimated using 5-fold cross-validation, which supports robust hyperparameter selection and generalisation control (Hastie et al., 2009). Model interpretation was supported using feature importance rankings and Partial Dependence Plots (PDPs), which allowed for evaluation of predictor relevance and marginal effects. These tools are discussed in greater detail in Section 4.2.6.

Together, these modelling choices support a predictive framework that is both accurate and interpretable. The following sections describe the modelling environment, training process, validation approach, model evaluation metrics, robustness checks, and interpretive techniques in greater detail.

4.2.1 Software and Tools

The modeling workflow was developed and executed in Python. Key libraries included Pandas and Numpy for data manipulation, Scikit-learn for RF modeling, cross-validation, performance metrics, and PDP generation, Optuna for parameter tuning and Matplotlib for static visualisation (Table 2). To ensure consistency in model training and evaluation, random processes were controlled using a fixed random seed (random_state = 42), which helps ensure that results are reproducible across multiple runs. The technical setup ensured a consistent modelling environment for training and validation, supporting stable and reproducible model results.

Table 2: Model components and key functions

Component	Details
Programming Language	Python
Key Libraries	pandas, numpy, scikit-learn, optuna, matplotlib
Function of Libraries	Data manipulation (pandas, numpy); RF modeling, validation, metrics, PDPs (scikit-learn); visualization (matplotlib)
Reproducibility Setup	Fixed random seed: random_state = 42

4.2.2 Model Training

The full constructed dataset was used in a 5-fold cross-validation. In each fold, the observations were randomly divided so that 80% formed the training set and 20% formed the testing set. This approach replaced a single, fixed 80/20 split—because a separate hold-out set yielded unstable results given the limited data—and ensured that every observation appeared in a test set exactly once across the five folds. Since both outcome variables are continuous, a simple random partition within each fold was appropriate. Stratification which is typically used to preserve class distributions in classification tasks, was not necessary in this regression setting.

Feature selection was handled internally by the Random Forest algorithm. All variables were included in the model without manual pre-filtering, allowing the model to empirically evaluate each feature's relevance. Although Random Forest is robust to multicollinearity, the initial variable set was theory-informed, limited to variables that are conceptually meaningful, publicly available and consistent across both time periods.

Hyperparameter tuning for the Random Forest regressions was performed using Optuna, which runs sequential hyperparameter trials, using each trial's results to guide the next within predefined ranges (Table 3) to identify the best performing parameter set.

Table 3: Summary table of model hyperparameters and testing values

Hyperparameter	Ranges	Description
n_estimators	50-300	Number of trees in the forest
max_depth	None, 5-50	Maximum depth of each tree
min_samples_leaf	1-20	Minimum number of samples required to be at a leaf node
max_features	'sqrt', 'log2', 'none'	Number of features considered at each split
Cross-validation	5-fold	Used to evaluate combinations during parameter tuning

4.2.3 Validation

Model validation employed two complementary methods. First, out-of-bag (OOB) error estimation provided an internal check by evaluating predictions on observations excluded from each tree's bootstrap sample (Breiman, 2001). Second, 5-fold cross-validation served as an external validation approach, with each fold using 80 percent of data for training and 20 percent for testing, and performance metrics aggregated across folds (Hastie et al., 2009). This combination ensured both internal (OOB) and external (cross-validation) measures of predictive performance.

4.2.4 Evaluation Metrics

Model performance was assessed separately for the two target variables. To evaluate predictive accuracy, two standard metrics were used. R^2 (R-squared) indicates the proportion of variation in the observed outcomes that is captured by the model's predictions, with higher values reflecting stronger predictive performance. Root Mean Squared Error (RMSE) was used to measure the average size of the prediction errors, reported in the same units as the outcome variable for ease of interpretation. R^2 offers insight into overall model explanatory power, while RMSE provides an interpretable measure of predictive accuracy (Hastie et al., 2009; Zheng, 2023). Both R^2 and RMSE were averaged across a 5-fold cross-validation procedure to ensure robust estimates of predictive performance.

Following Shmueli's (2010) distinction between explanatory and predictive modelling, the evaluation focuses on predictive performance and the practical relevance of model outputs rather than on causal inference.

4.2.5 Robustness Check

To assess how predictive performance varies with sample composition, subsamples were created for each outcome variable using two segmentation approaches: population thresholds and municipality types. In the first approach, municipalities below 10,000 and 20,000 inhabitants were excluded to focus on structurally larger municipalities (Table 4). The 10,000 threshold aligns with the SCIRN definition of shrinking cities (Wiechmann, 2008). The 20,000 threshold was introduced to test whether performance continues to improve as smaller units are progressively removed.

Table 4: Description of the full sample and population thresholds.

Sample	Description
Full sample	Includes all 290 municipalities

Population Thresholds	Description
Total Population (TP) > 10,000	Isolates smaller municipalities more prone to shrinkage, following the threshold used by Peng et al. (2023) for identifying structural vulnerabilities.
Total Population (TP) > 20,000	Excludes even more small municipalities to test whether predictive performance improves with progressively larger population bases.

The second approach uses a municipality-type classification adapted from Boverket's national typology which combines population size with functional characteristics such as commuting patterns. For instance, whereas Boverket distinguishes between metropolitan cities (Storstäder) and municipalities within commuting distance to metropolitan cities (Pendlingskommun nära Storstäder), this study merges both into a single "Metropolitan" group. In total, municipalities are divided into four broader categories: (1) Metropolitan municipalities, (2) Medium and large city areas, (3) Smaller cities and towns, and (4) Rural municipalities (Table 5) (Boverket, 2024). For the robustness check, municipalities were progressively excluded by type: first all type 4 were removed, and then type 3 municipalities were also excluded.

Table 5: Description of municipality-type classifications.

Municipality-Type Classification	
1 - Metropolitan municipalities	Includes the three largest cities: Stockholm, Gothenburg, Malmö and municipalities functionally connected to the metropolitan cores via commuting. High population density, complex infrastructure, strong economic centres.
2 – Medium/large city area	Regional hubs with a broad labour market and administrative services. Independent economic function but not part of the three big metro regions. Also including suburbs or satellites of the larger city.
3 - Smaller cities and towns	Local urban centres with limited but significant regional influence.
4 - Rural municipalities	Low-density areas with dispersed settlement structures. Often face depopulation and service accessibility challenges.

For each subsample, the full RF modelling procedure was re-executed. The results of these robustness checks guided the choice of a sample definition for subsequent feature-importance and PDP analyses.

4.2.6 Model interpretation

Feature-importance was computed via permutation-based methods, which assess how much model error increases when a single variable is randomly permuted (Hastie et al., 2009; Molnar, 2019). Using Scikit-learn's implementation on OOB data, each feature is permuted in turn while all others remain constant at their mean value, and the resulting rise in mean squared error (MSE) is recorded. The relative increase in MSE then serves as the importance score for that variable.

To examine the form of key predictor–outcome relationships, Partial Dependence Plots (PDPs) were generated for the highest-importance variables. PDPs display how predicted values change as one variable varies while holding all others constant at their mean. Moreover, PDPs can reveal nonlinear patterns—particularly stepped effects—which, when they appear consistently, indicate threshold points where predictive relationships shift abruptly at specific variable values (Hastie et al., 2009; James et al., 2013; Molnar, 2019). While alternative interpretability techniques, such as SHAP values or Accumulated Local Effects (ALE), could offer additional nuance, PDPs were chosen for their simplicity (Molnar, 2019).

4.3 Limitations and Ethical Considerations

4.3.1 Data Limitations

Overall data quality was high, with core variables sourced from standardized national databases. During data exploration, SCB data emerged as the most practical foundation for modeling due to its national coverage and consistent formatting across variables. However, the key rent-per-square-meter data were only available starting in 2016, which constrained the temporal scope and led to an iterative exploratory approach that identified four-year intervals as optimal—hence the overlapping 2016–2020 and 2019–2023 cross-sectional periods.

Several potentially useful built-environment indicators—such as building permits, vacancy rates, and commercial real estate data—were collected by SCB on either regional level or by metropolitan areas, but unavailable at the municipal-level. The only publicly accessible house-price data on the municipal-level pertained to single-family homes. This limitation may influence results in larger municipalities, where multifamily markets are more prevalent. Private-sector datasets (e.g., transaction-level price data or detailed permit records) offered finer geographic granularity but proved incompatible with SCB's formatting or unreliable for smaller municipalities. Integrating these sources would require extensive cleaning, matching, and variable construction, resources beyond this study's scope. Consequently, proxy variables were constructed to capture housing-market dynamics pragmatically, acknowledging that such proxies may not fully reflect local conditions.

4.3.2 Methodological Limitations

Feature-importance scores indicate predictive contribution but not causal influence (Breiman, 2001; Shmueli, 2010). Predictive models are constrained by the distribution and patterns present in the training data (Hastie et al., 2009) and cannot anticipate sudden structural changes such as sudden economic shifts, demographic shocks, or unexpected policy reforms.

Although data-driven methods offer valuable insights, they are shaped by conceptual and institutional frameworks and must be interpreted accordingly (Shmueli, 2010; Molnar, 2019). Predictive tools, in particular, should be viewed within their broader context to avoid overly definitive or simplified conclusions (Kitchin, 2014).

4.3.3 Ethical Considerations

This study uses publicly available data and therefore presents no risks related to individual privacy or personal data protection. However, predictive modelling of municipal shrinkage has ethical considerations beyond technical implementation. Municipalities identified as “at risk” may face reputational damage, investment hesitation, or reduced institutional confidence, particularly in rural or already vulnerable contexts. This study does not classify or label individual municipalities, but instead models continuous variation in population and housing price change to identify structural patterns. However, predictive tools are increasingly used in risk scoring or binary classification applications, which can lead to unintended social or political consequences if not framed carefully (Kitchin, 2014).

To mitigate these risks, model outputs are explicitly framed as probabilistic risk assessments rather than deterministic forecasts. Interpretations are qualified with uncertainty and designed to support, not replace, qualitative local knowledge and contextual judgment (Shmueli, 2010; Molnar, 2019).

5 Results

Chapter 5 begins by describing municipal population and housing-price dynamics in Sweden, reporting the share of municipalities experiencing growth or decline between 2016–2020 and 2019–2023 and illustrating shifts in the distribution of Population Change Rate (PCR) and House-price Change (HPC) across all Municipality Types (MT) and in total (Tables 6–7; Figs. 1–2). It then presents Random Forest (RF) regression results, comparing model performance for PCR and HPC across the full sample and various sample cutoffs via R^2 and RMSE metrics (Table 8; Fig. 3). Next, feature-importance scores are detailed for both models on the Total Population (TP) > 10,000 sample, highlighting the top predictors of PCR and HPC (Tables 9–10). Finally, partial-dependence plots (PDPs) reveal nonlinear and threshold effects, unpacking how key demographic, economic, and housing-market indicators influence municipal outcomes in each period (Figs. 4–5).

5.1 Population and Housing Price Dynamics in Sweden

Table 6 summarizes the number of growing and shrinking municipalities during each period, broken down by MT and for the full sample of 290 municipalities. Between 2016 and 2020, 205 municipalities (70.7 %) experienced population growth, while 85 (29.3 %) faced net population decline. In the subsequent period, 2019 to 2023, the number of growing municipalities decreased to 153 (52.8 %), while those undergoing shrinkage rose to 137 (47.2 %). Table 7 offers a longitudinal breakdown, showing how municipalities followed different outcome trajectories over the full 2016–2023 period. A total of 148 municipalities (51.0 %) sustained continuous growth, 80 (27.6 %) experienced continuous shrinkage, 57 (19.7 %) shifted from growth to shrinkage, and only 5 (1.7 %) reversed from shrinkage to growth.

Table 6: Shrinking and growing municipalities by municipality-type classification and timeframe.

	2016-2020				**2019-2023**			
Municipality Type Classification	**Growing**		**Shrinking**		**Growing**		**Shrinking**	
	Count	**(%)**	**Count**	**(%)**	**Count**	**(%)**	**Count**	**(%)**
1 - Metropolitan Municipalities	46	100	0	0	42	91.3	4	8.7
2 - Larger Cities	86	79.6	22	20.4	66	61.1	42	39
3 - Smaller Cities and Towns	56	69.1	25	30.9	34	42	47	58
4 - Rural Municipalities	17	30.9	38	69.1	11	20	44	80
Total	**205**	**70.7**	**85**	**29.3**	**153**	**52.8**	**137**	**47.2**

Table 7 : Population trends grouped by municipality-type classifications, full sample between 2016-2023

Municipality Types	Continuous Growth		Continuous Shrinkage		Growth to Shrinkage		Shrinkage to Growth	
	Count	(%)	Count	(%)	Count	(%)	Count	(%)
1 - Metropolitan Municipalities	42	91.3	0	0	4	8.7	0	0
2 - Larger Cities	65	60.2	21	19.4	21	19.4	1	0.9
3 - Smaller Cities and Towns	32	39.5	23	28.4	24	29.6	2	2.5
4 - Rural Municipalities	9	16.4	36	65.5	8	14.5	2	3.6
Total	**148**	**51.0**	**80**	**27.6**	**57**	**19.7**	**5**	**1.7**

Examining these outcomes by MT reveals pronounced structural disparities between urban and rural areas. Metropolitan municipalities show the greatest stability, with 42 out of 46 municipalities sustaining continuous growth, with the remaining four experiencing growth first and then shrinkage. Large city areas display more mixed patterns as 60% of the municipalities maintain growth, while roughly 40% experience shrinkage in the second period. This pattern reverses in small cities, while rural municipalities appear most vulnerable with only 11 (20%) experiencing growth in the second period.

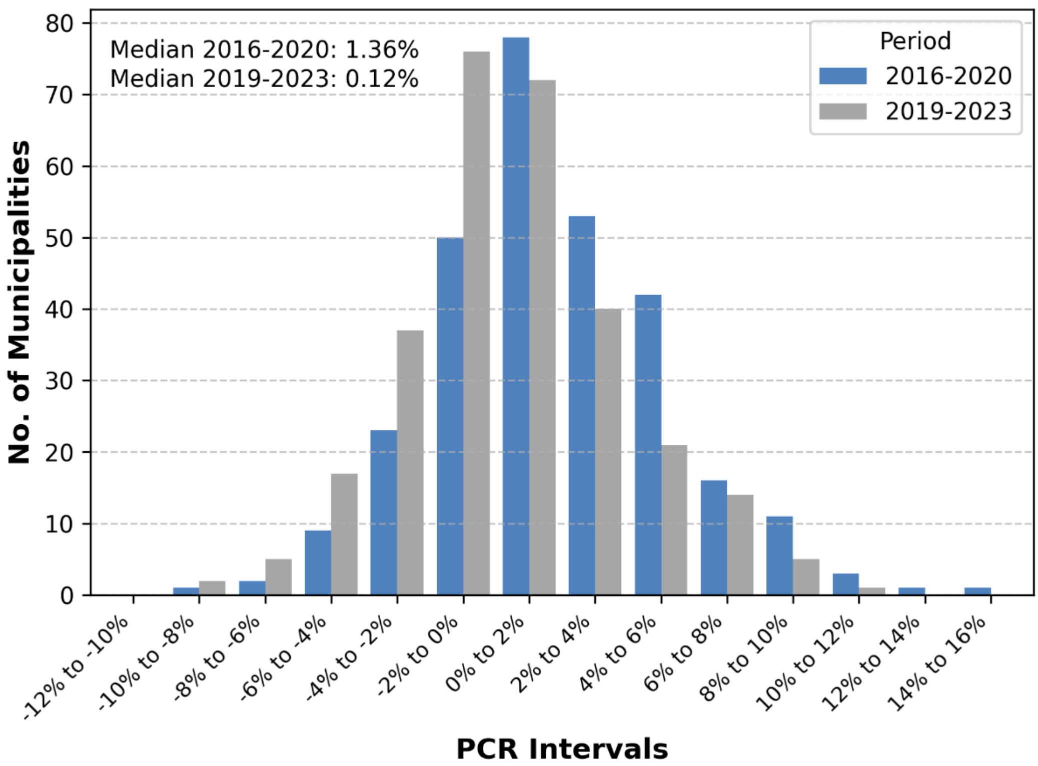


Figure 1: Distribution of PCR in the two study periods.

Fig. 1 illustrates the shifts in PCR, by showing how municipalities are distributed across 2-percentage-point intervals, comparing the two time periods side by side. Fig.

2 illustrates the distribution in 5-percentage-point intervals of HPC across municipalities for the two periods. Between 2016 and 2020, house-price growth was both stronger and more heterogeneous, with a median HPC of 24.2%. In contrast the 2019 to 2023 period saw more moderate and uniform price increases, with a lower median HPC of 17.2%. This downward shift in both magnitude and variability mirrors the parallel decline observed in PCR across municipalities.

The HPC distribution for 2016–2020 exhibits substantial dispersion around the median, with both moderate and extreme growth rates relatively evenly represented (Fig. 2). This produces a flattened central peak and extended tails on both sides. In contrast, the 2019–2023 distribution is more compressed, with the majority of municipalities concentrated in the 10% to 20% range. Few municipalities experienced either sharp declines or exceptionally large increases, resulting in a more pronounced peak and noticeably thinner tails.

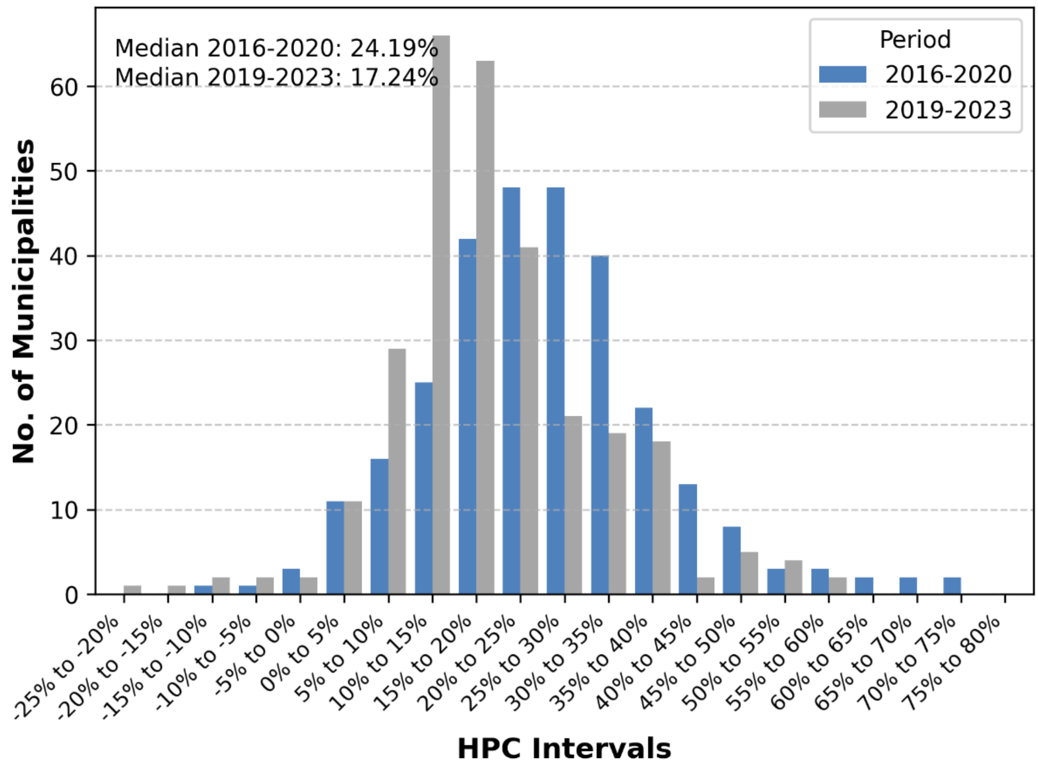


Figure 2: Distribution of HPC in the two study periods.

Overall, the second period reflects lower variability in HPC and a clear shift toward more uniform, moderate increases. The downward shift in HPC between periods mirrors the simultaneous downward shift in the PCR distribution across municipalities (Fig. 1), highlighting that both outcome variables followed a trajectory of contraction between periods.

5.2 Random Forest Regression Results

5.2.1 Model Performance Results

RF regression was used to predict both PCR and HPC across the two periods (2016-2020 and 2019-2023), based on the five sample definitions outlined in section 4.2.5. Results from the 5-fold cross-validation of the RF models (Table 8; Fig. 3) show that the PCR model consistently outperformed the HPC model. R^2 values for PCR range from 0.44 to 0.67 across all samples, substantially higher than those for HPC, which range from 0.10 to 0.41. Similarly, the PCR model yielded lower and more tightly clustered RMSE values (0.017–0.022), compared to the broader range observed in the HPC model (0.062–0.117), indicating greater predictive precision for population change than for housing-price movements.

As shown in Table 8, the PCR model achieved its highest performance when applied to the full sample of 290 municipalities, with R^2 values of 0.67 and 0.65 for the respective periods. Applying sample cutoffs, either by TP or by MT, resulted in reduced R^2 values, while RMSE remained stable at approximately 0.02 across all sample definitions. This suggests that the excluded data contained meaningful predictive information for PCR.

Table 8: Model performance across sample cutoffs.

		PCR model				HPC model			
Sample definition	**No. of observations**	**2016–2020**		**2019–2023**		**2016–2020**		**2019–2023**	
		R^2	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**
Full sample	290	0.67	0.020	0.65	0.020	0.23	0.114	0.15	0.117
TP < 10,000	218	0.62	0.020	0.55	0.020	0.33	0.085	0.11	0.088
TP < 20,000	123	0.65	0.017	0.44	0.020	0.35	0.071	0.12	0.062
MT [1, 2, 3]	235	0.55	0.021	0.51	0.021	0.40	0.092	0.10	0.107
MT [1, 2]	154	0.56	0.021	0.49	0.022	0.41	0.088	0.12	0.107

Although overall model performance was modest for HPC, results were notably stronger in the 2016–2020 period (R^2 = 0.23–0.41) compared to 2019–2023 (R^2 = 0.10–0.15). Within the 2016-2020 window, restricting the sample by TP or MT improved model performance (R^2 = 0.33–0.41; RMSE = 0.07–0.09) relative to the full sample (R^2 = 0.23; RMSE = 0.11), indicating that lower-population and rural municipalities introduced additional noise. However, when applying stricter cutoffs, MT [1, 2] or TP > 20,000, R^2 values remained largely unchanged, but model stability declined sharply due to reduced sample size. This highlights the trade-off between filtering out noisy observations and maintaining a robust number of observations, with more inclusive thresholds (MT [1, 2, 3]; TP > 10 000) emerging as the preferable balance.

Restricting to MT [1, 2, 3] yielded a marginally higher R^2 (0.40) than applying a TP > 10,000 cutoff (R^2 = 0.33). This difference may reflect greater noise in fully rural municipalities versus lower-populated municipalities with mixed urban characteristics, or it may simply result from the larger sample size for MT [1, 2, 3] (n = 235) compared to TP > 10,000 (n = 218).

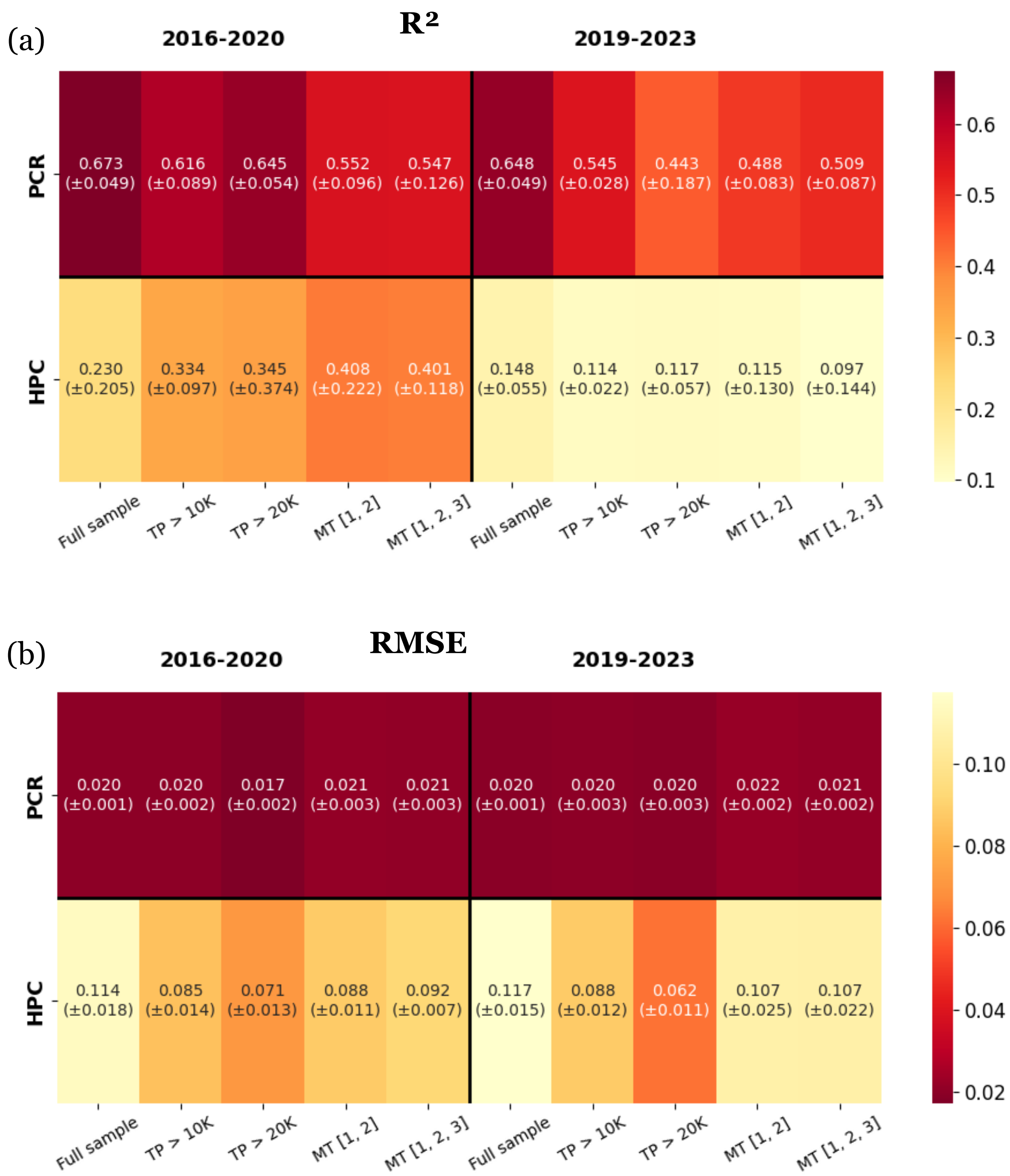


Figure 3: Heat maps of the average R^2 (a) and RMSE (b) from 5-fold cross validation of RF regression models predicting PCR and HPC across the five sample definitions and two periods.

Cross-validation (CV) results confirm that the RF model effectively captures cross-municipality variance in PCR, showing strong performance with low and consistent error. In contrast, the HPC model exhibited only modest explanatory power in the 2016–2020 period, even after applying sample restrictions. Since PCR performance was slightly stronger with the TP > 10,000 cutoff than with MT [1, 2, 3], and both cutoffs yielded similar performance for HPC, TP > 10,000 was selected as the common sample definition to support comparability across models. Based on the common sample definition, feature importance scores were calculated for all demographic, economic, and housing-market indicators, followed by the generation of partial dependence plots to explore the relationships between key predictors and the outcome variables.

5.2.2 Feature-Importance Scores

Permutation importance (Section 4.2.6) was used to estimate relative feature-importance for both RF models on the TP > 10,000 sample. Table 9 shows the importance scores and rankings for the PCR models variables. Seven features exceed 5 percent importance: RTP, ODR, DSC_lag3, PD, INC, AHP, and UPR, indicating their strong predictive influence on municipal population change, while all other variables exhibit relatively low importance.

Table 9: Feature-importance scores and rankings of variables in the PCR model.

PCR - Model (TP > 10000)	**2016 - 2020**		**2019 - 2023**	
Variables	**FI (%)**	**Ranking**	**FI (%)**	**Ranking**
RTP	20.95	1	16.95	1
ODR	16.11	2	11.83	3
DSC_lag3	13.69	3	15.37	2
PD	8.53	4	5.93	5
INC	7.05	5	11.6	4
AHP	5.72	6	3.34	13
UPR	5.68	7	3.85	7
TP	3.02	8	3.62	11
ALE	2.95	9	4.18	6
ER	2.82	10	3.85	8
MS	2.69	11	3.8	9
MR	2.59	12	2.76	16
INC_lag3	2.58	13	2.82	14
FPR	2.41	14	3.6	12
GDP_pc	1.82	15	2.77	15
HPC_lag3	1.39	16	3.75	10

In both the 2016-2020 and 2019-2023 period, RTP ranked first among all variables, accounting for 20.95% and 16.95% of the model's predictive power, respectively. This suggests that population shrinkage is strongly associated with high rent levels relative to local house prices. Between periods, the ranking of DSC_lag3 rose from third (13.69%) to second place (15.37%), swapping places with ODR, which declined from 16.11% to 11.83%. This shift suggests that changes in housing supply contributed more to the model's predictive power in the later period, while the influence of municipal demographic structure declined. However, it remains unclear whether this reflects genuine shifts in underlying relationships or interactions with other variables. Nevertheless, both variables remained relatively strong predictors of PCR across both periods. PD and INC completed the top five predictors, with PD declining from 8.53% to 5.93% and disposable income rising from 7.05% to 11.60%. The importance scores of AHP (5.72%) and UPR (5.68%) further underscore the predictive influence of demographic structures on future population change. These shifts between periods illustrate how feature importance scores reflect changes in the relative contribution of variables within the model over time. However, they do not, in themselves, provide evidence of causal change in underlying municipal dynamics (Table 9).

Table 10: Feature importance scores and rankings of variables in the HPC model.

HPC - Model (TP > 10000)	**2016 - 2020**		**2019 - 2023**	
Variable	**FI (%)**	**Ranking**	**FI (%)**	**Ranking**
INC	28.48	1	5.02	12
HPC_lag3	21.59	2	12.02	3
ODR	7.17	3	8.99	5
TP	7.14	4	5.07	11
MS	6.64	5	5.23	10
SD_1000	5.6	6	6.41	6
FPR	4.34	7	13.88	1
PCR_lag5	4.21	8	9.59	4
PD	3.79	9	12.49	2
GDP_pc	3.18	10	3.76	13
INC_lag3	2.69	11	5.53	9
ER	2.68	12	6.38	7
ALE	2.5	13	5.62	8

Table 10 shows the feature importance scores and rankings for the HPC model. As mentioned in Section 5.2.1, the HPC model performance varied considerably between the two periods, with R^2 falling from 0.33 in 2016–2020 to just 0.09 in 2019–2023. When explanatory power is low, as observed in the second period (Section 5.2.1), it indicates that none of the predictor variables capture the underlying signal effectively. This typically leads to a flattened importance distribution, where most variables contribute only marginally and to a similar extent, an effect visible in Table 10. In contrast, well-performing models tend to exhibit a sharper distinction between high- and low-importance variables, as seen in the PCR model (Table 9). For this reason, the feature importance analysis of the HPC model will focus primarily on the 2016–2020 period, where the results are more robust and interpretable.

In this period, 6 variables exceed 5% importance: INC, HPC_lag3, ODR, TP, MS, and SD_1000. INC (28.48%) and HPC_lag3 (21.59%) separate themselves at the top, suggesting that house price changes in Swedish municipalities are most influenced by their respective income levels and local housing market momentum. ODR (7.17%) and TP (7.14%) rank third and fourth, underscoring the relevance of demographic structure to housing-price dynamics. FPR, PCR_lag5, and PD, each contribute roughly 4%, bringing the combined importance of demographic variables to over 29%.

5.2.3 Partial Dependence Analysis

For the TP > 10,000 sample, PDPs were used to examine the nature of the relationships, whether positive or negative, linear or non-linear, between each variable and the outcome variables (Figs. 4–5). The Random Forest curves display stepped, non-linear patterns and reveal threshold effects for most variables, as anticipated in Section 4.2.6. Key patterns for the top predictors in each model are summarized below with a reference to the corresponding figure.

Fig. 4 reveals a strong, nonlinear negative relationship between RTP and PCR for both 2016–2020 and 2019–2023. When the RTP ratio remains below approximately 0.35, PCR stays relatively high (around 4% in 2016–2020 and 2% in 2019–2023), indicating that municipalities where rent costs are low relative to house prices experience stronger population growth or retention. As the RTP ratio increases from 0.35 to 0.55, PCR declines sharply, after which the slope of the curve tapers off. This steep decline appears at consistent RTP levels for both periods, underscoring a critical affordability threshold above which rental-cost burdens become a major driver of shrinkage.

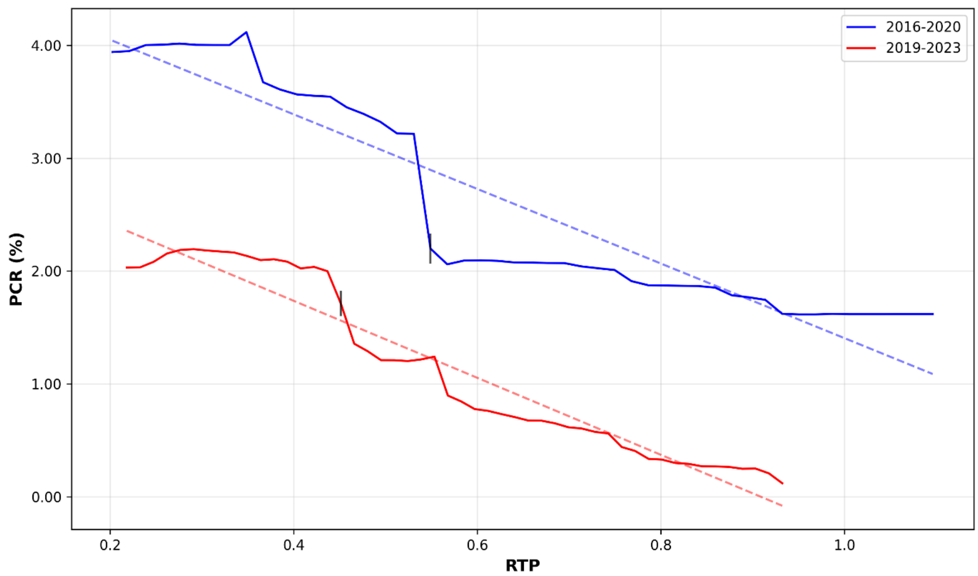


Figure 4: PDP showing correlations between RTP and PCR; solid curves refer to partial dependence interpreted by PCR model, and dotted lines are the linear trendlines of their corresponding curves.

Fig. 5 shows that in 2016–2020, PCR remains stable near 4% while ODR is below 0.35, then declines steeply to about 2% as ODR approaches 0.40. Similarly, in 2019–2023, PCR falls from roughly 2.4% to 1.0% over the same ODR range. These inflection points at, ODR ≈ 0.35 and ODR ≈ 0.40, indicates that crossing these aging thresholds accelerates shrinkage, highlighting the impact of an increasing elderly share on population dynamics.

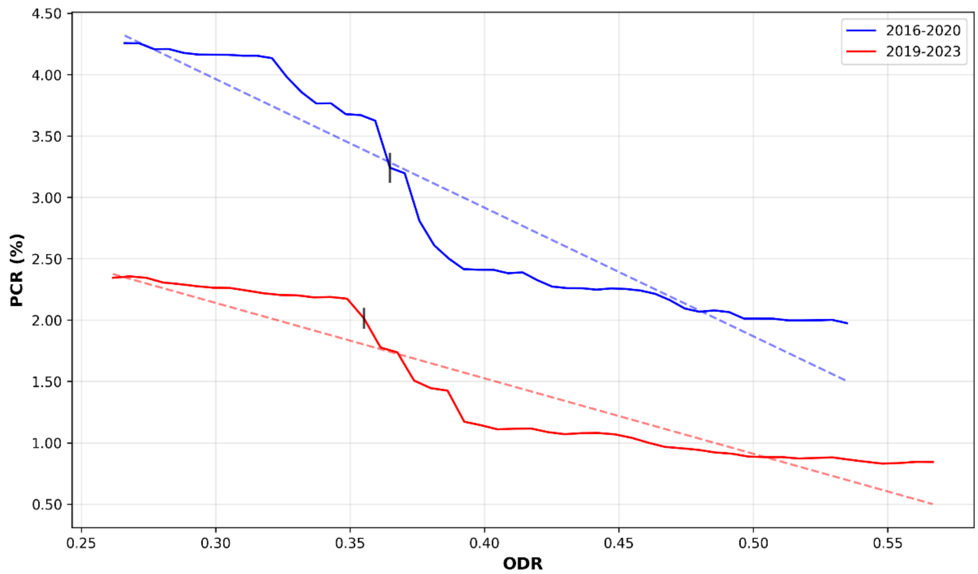


Figure 5: PDP showing correlations between ODR and PCR; solid curves refer to partial dependence interpreted by PCR model, and dotted lines are the linear trendlines of their corresponding curves.

Fig. 6 illustrates that in 2016–2020, PCR increases nearly linearly with DSC_lag3, rising from roughly 2.2% when DSC_lag3 is below 10 additional dwellings per 1,000 inhabitants to about 4% when SD_1000 exceeds 25, with a slight nonlinear inflection around DSC_lag3 ≈ 17. In 2019–2023, the relationship becomes more stepped. PCR climbs until it levels off at 6% once DSC_lag3 surpasses 5, then follows a steeper rise between 13 and 22 before levelling off near 2.5%. This added nonlinearity likely explains the observed increase in the feature-importance scores of DSC_lag3 between periods, as stronger threshold effects enhance its predictive contribution.

Fig. 7 shows that in 2016–2020, PCR jumps from 2.3% at PD below 50 inhabitants/km² to 3.3% at densities above 100 inhabitants/km², after which it remains flat. In 2019–2023, PCR likewise rises from 1.2% to 1.7% around the same density range. These patterns indicate a threshold near 100 inhabitants/km², below which sparsely populated municipalities are most vulnerable to shrinkage, while further densification yields diminishing returns for population growth.

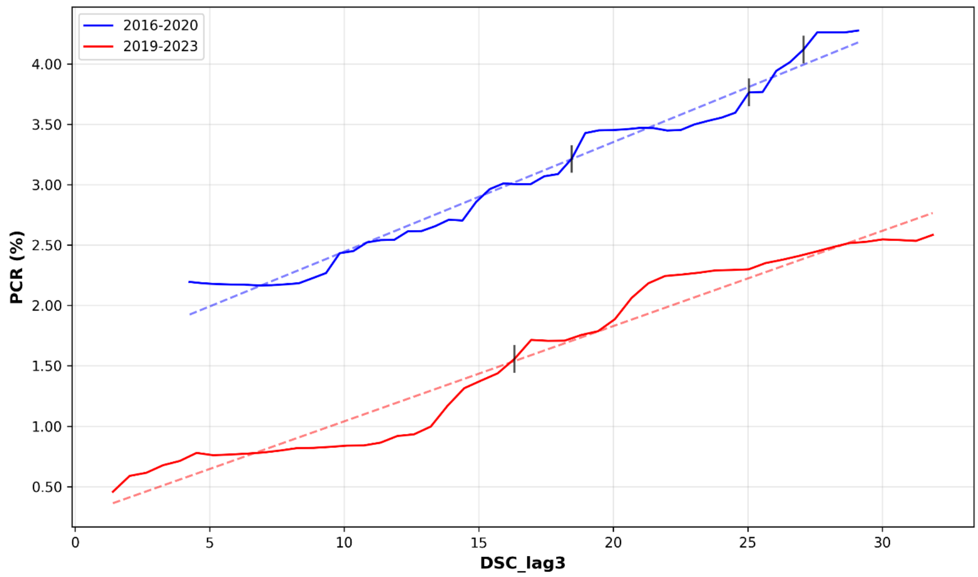


Figure 6: PDP showing correlations between DSC_lag3 and PCR; solid curves refer to partial dependence interpreted by PCR model, and dotted lines are the linear trendlines of their corresponding curves.

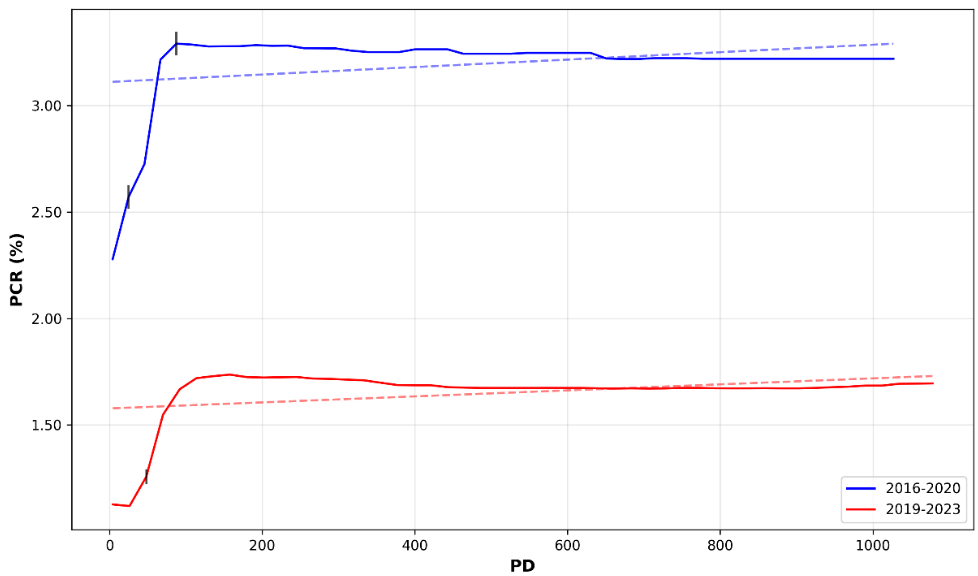


Figure 7: PDP showing correlations between PD and PCR; solid curves refer to partial dependence interpreted by PCR model, and dotted lines are the linear trendlines of their corresponding curves.

Fig. 8 shows that in 2016–2020, predicted PCR rises from approximately 2.4% at a disposable income level (INC) of SEK 240,000 per capita to 2.7% at SEK 270,000, then increases more steeply between SEK 270,000 and SEK 280,000 before levelling off. In 2019–2023, the curve follows a similar shape but is shifted rightward—reflecting higher national income levels—so that predicted PCR climbs from roughly 0.6% to 2.0% before reaching SEK 290,000 and then flattens. This pattern indicates that beyond these income thresholds, additional income yields no further population growth—though this flatness may reflect few observations past the threshold—and that PCR was more responsive to lower income levels in the second period.

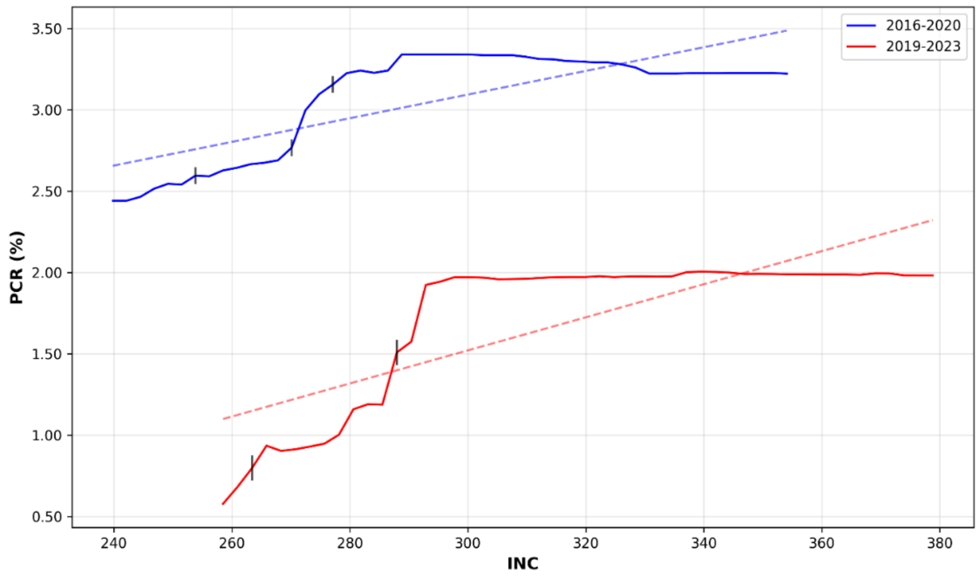


Figure 8: PDP showing correlations between INC and PCR; solid curves refer to partial dependence interpreted by PCR model, and dotted lines are the linear trendlines of their corresponding curves.

Fig. 9 reveals a delayed nonlinear inflection in the relationship between PCR and TP. In 2016–2020, predicted PCR steadily increases until TP reaches roughly 50,000, after which it levels off; in 2019–2023, that inflection point shifts to about 70,000 before flattening. Both curves exhibit a slight negative slope between TP ≈ 10,000 and 15,000, likely reflecting a noisy outlier in the sample set. Modest positive correlations reinforce the feature-importance findings (Section 5.2.2), demonstrating that TP plays a secondary role compared to predictors like HSC_lag3 or RTP.

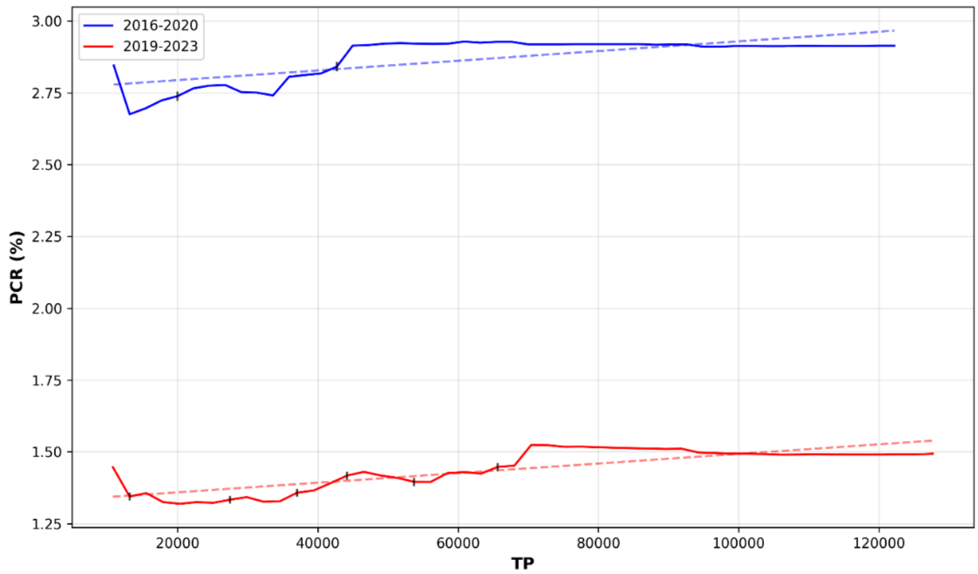


Figure 9: PDP showing correlations between TP and PCR; solid curves refer to partial dependence interpreted by PCR model, and dotted lines are the linear trendlines of their corresponding curves.

Figs. 10-13 show PDPs for the HPC model across periods. In 2016–2020, predictors display clear threshold effects, while in 2019–2023 they produce much flatter, noisier curves. In the first period, variables like INC, HPC_lag3, and ODR exhibit distinct inflection points while in the second period, these marginal effects mostly disappear. As a result, the following analysis will concentrate on the 2016–2020 PDPs and their interpretable thresholds.

Fig. 10 reveals a negative, nonlinear correlation between HPC and INC, likely reflecting greater market maturity and price stability in higher income municipalities. In 2016–2020, predicted growth declines from about 28% to 19% in steep steps at INC thresholds near SEK 245,000, SEK 265,000, and SEK 280,000 before levelling off at SEK 290,000. In contrast, the 2019–2023 curve remains flat at roughly 18% growth, indicating that disposable income provided minimal predictive signal during that period.

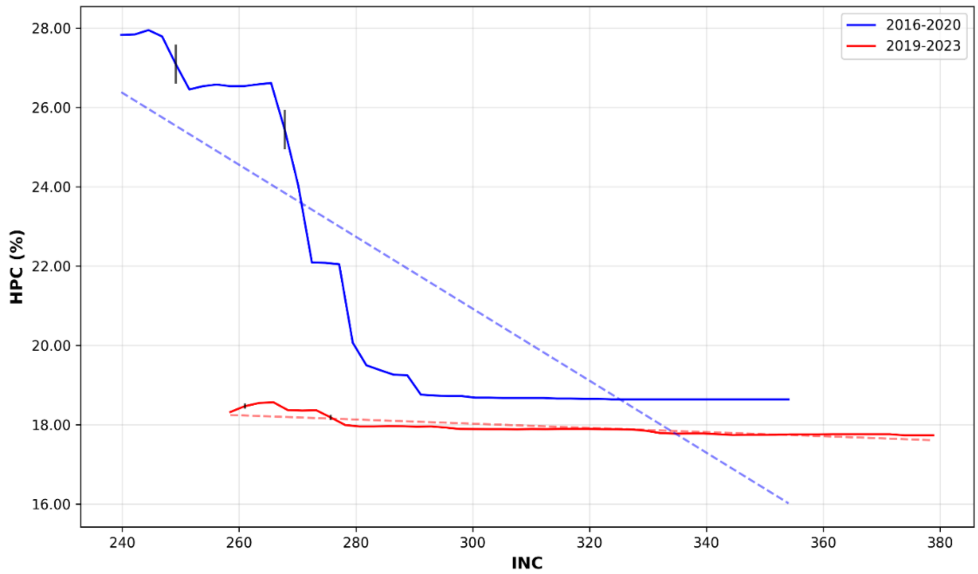


Figure 10: PDP showing correlations between INC and HPC; solid curves refer to partial dependence interpreted by PCR model, and dotted lines are the linear trendlines of their corresponding curves.

Fig. 11 shows a negative correlation between HPC and prior house-price momentum (HPC_lag3) and subsequent HPC, with the PDP closely following a linear decline in both periods and lacking clear threshold effects. This suggests that municipalities with stronger recent price gains tend to see smaller following increases, and vice versa. Although the 2019–2023 curve is somewhat flatter—indicating a weaker signal—the indicator still contributes meaningfully to explanatory power in both periods.

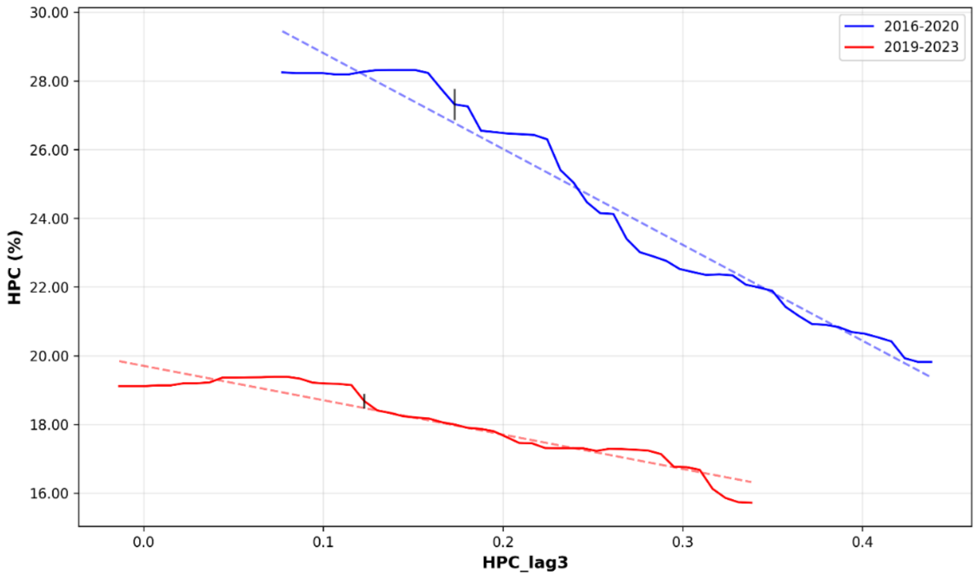


Figure 11: PDP showing correlations between HPC_lag3 and HPC; solid curves refer to partial dependence interpreted by PCR model, and dotted lines are the linear trendlines of their corresponding curves.

Fig. 12 shows that in 2016–2020, predicted HPC remains stable at approximately 22.5% for ODR values below 0.28, then rises sharply to about 24.5% as ODR approaches 0.36, and gradually declines to 24% at higher ODR values. This nonlinear pattern indicates a strong positive effect of increasing elderly share on house-price growth within the 0.28–0.36 range, followed by a subtle negative drift beyond that peak, suggesting diminishing returns at very high dependency levels.

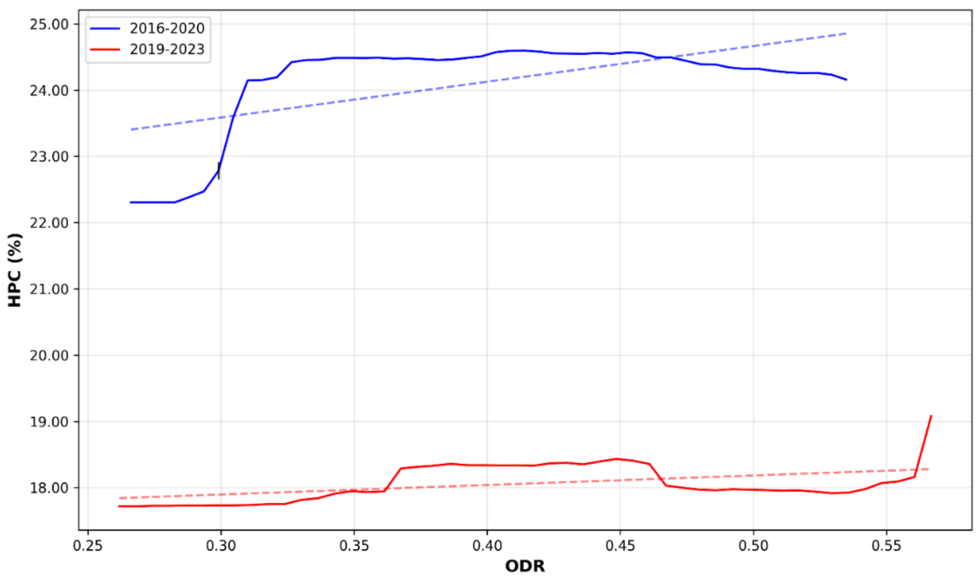


Figure 12: PDP showing correlations between ODR and HPC; solid curves refer to partial dependence interpreted by PCR model, and dotted lines are the linear trendlines of their corresponding curves.

Figure 13 reveals a nonlinear, stepped decline in housing-price growth as TP increases. In 2016–2020, predicted HPC peaks at roughly 26% for municipalities with TP ≈ 10,000, then falls to about 24% as TP rises to ≈ 19,000. A second, more gradual drop occurs between TP ≈ 35,000 and ≈ 50,000, after which growth stabilizes. These inflection points indicate that initial increases in population size strongly suppress price growth up to those thresholds, with further size increases exerting little additional impact. Note that this relationship may be influenced by the nature of the underlying data, which covers only single-family housing prices. In larger municipalities, markets for single-family homes are typically more mature than those for multifamily units, potentially dampening observable effects at higher population sizes.

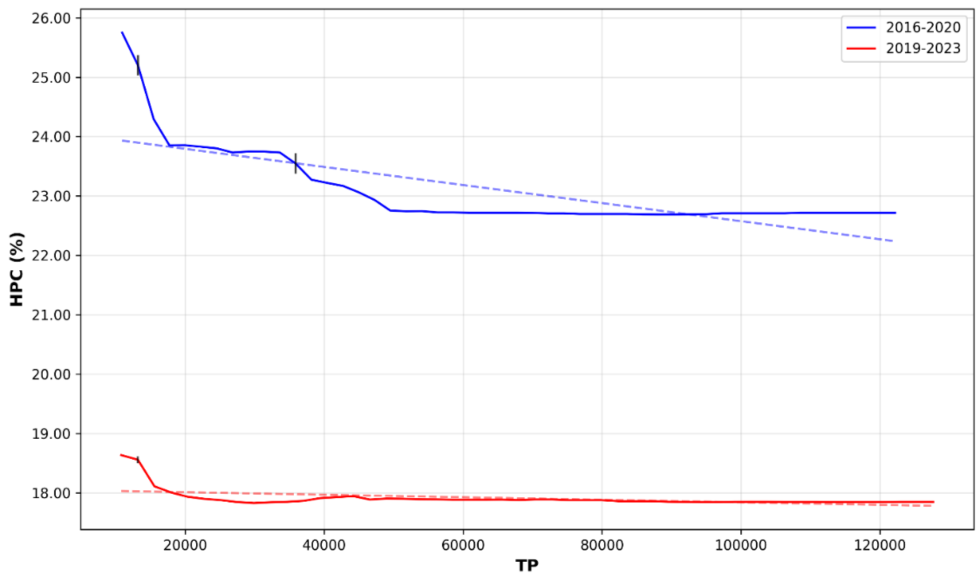


Figure 13: PDP showing correlations between TP and HPC; solid curves refer to partial dependence interpreted by PCR model, and dotted lines are the linear trendlines of their corresponding curves.

Overall, the PDPs reveal several non-linear and threshold-based relationships between predictors and outcomes. For PCR, demographic and spatial variables such as population density, dependency ratios, and housing-stock changes show distinct inflection points, indicating that their influence varies across value ranges. For HPC, threshold effects are present in the 2016–2020 period but largely disappear in 2019–2023, with flatter and more variable patterns reflecting reduced model performance. These findings suggest that the predictive relationships identified by the RF model are stronger and more consistent for population change than for housing-price dynamics, particularly in the second period.

6 Discussion

6.1 Shrinkage Effects in Swedish Municipalities

Sweden's municipal landscape between 2016–2020 and 2019–2023 illustrates a clear shift from widespread growth toward a near equilibrium of expansion and contraction, signaling that the demographic, economic and housing-market pressures driving shrinkage are intensifying. Larger urban centers, especially metropolitan areas, have largely maintained resilience, likely drawing on agglomeration benefits and diversified economies, but smaller cities and rural communities are increasingly vulnerable.

In the PCR model for both timeframes, the RTP ranks first in feature-importance, while DSC_lag3 appears among the top three predictors, rising between periods. This highlights that housing-market dynamics—such as rental affordability and the pace of new construction—are among the strongest drivers of municipal population change. Which aligns with prior research linking construction activity and housing mismatch to structural demographic decline (Lindh and Malmberg, 2008). Interpreting the PDPs through Cumulative Causation Theory (CCT), high rent burdens beyond a 0.35 RTP threshold (Fig. 4) act as a tipping point that accelerates out-migration by eroding local purchasing power and service funding, creating a self-reinforcing cycle of decline. This self-reinforcing effect aligns with previous findings that connect affordability pressures and housing imbalances to downward feedback loops in municipal viability (Haase et al., 2014). This steep drop at similar RTP levels across both periods underscores this critical affordability threshold, above which rental-cost burdens become a major driver of shrinkage.

The positive relationship between DSC_lag3 and PCR suggest that shrinking demand leads municipalities to underinvest in housing upkeep, reinforcing vacancies and further population loss. An outcome that is consistent with findings that highlight fiscal constraints and planning inertia in declining municipalities (Eimermann et al., 2022; Grundel and Magnusson, 2023). The relationship becomes notably more nonlinear in the latter period, with a steeper slope around thresholds. This indicates a stronger impact between those thresholds, echoing the rising feature-importance between periods as stronger threshold effects enhance its predictive contribution.

In the HPC model, INC and HPC_lag3 dominated importance scores, together accounting for half the model's explanatory power. Municipal income levels show a strong negative relationship with both PCR and HPC below a threshold of SEK 290,000, while additional income beyond the threshold exerts diminished impact on both effects. Prior movements in housing prices also show a negative relationship with

HPC but have limited signal on PCR. Interpreting through Urban Life Cycle Theory (ULCT), income-driven price growth signals markets in a growth or early maturity phase, whereas strong prior momentum can either bolster or temper further gains depending on local stock conditions. The findings reflect this maturity and saturation effects that ULCT predicts for housing and labour markets. The patterns is also consistent with prior findings that link income stratification and housing-market maturity to divergence in urban growth trajectories, particularly in later stages of the urban life cycle (Haase et al., 2014; Malmberg, 2010).

The positive relationship between TP and HPC below TP threshold of 50,000 indicates that initial population increases tend to dampen price growth, reflecting the greater market maturity of larger settlements. This pattern may partly stem from HPC only reflecting single-family housing prices, since in larger municipalities, single-family markets often reach maturity earlier than multifamily-markets, potentially dampening observable effects at higher population sizes.

Strong positive correlations of PCR with PD < 100 inhabitants/km², supported by a mild positive correlation with TP < 50,000 underline the vulnerability of small, sparsely settled municipalities. These scale effects resonate with the early phases of ULCT in which under-sized or low-density settlements lack the critical mass to sustain local services and markets, as noted in the case study of Åsele municipality (Eimermann et al., 2022). However, these small sparsely settled municipalities also exhibit higher variance and idiosyncratic behaviour, which introduces noise when pooled with larger, more stable municipalities. This was observed as model performance improved when introducing sample cutoffs in model testing, implying that rural municipalities and small populations include more noisy data that compromise predictive accuracy. This reflects broader challenges in modelling heterogeneous urban systems, where variance and spatial idiosyncrasies often affect generalisability (Breiman, 2001; Hastie et al., 2009).

ODR was an important predictor for both models. Its negative correlates with PCR suggests that an ageing population may contribute to population decline in Swedish municipalities, which is consistent with patterns observed in later stages of the Demographic Transition Theory (DTT), characterized by lower fertility, increasing life expectancy, and a tendency toward natural decline. Likewise, the critical thresholds identified for PD and TP aligns with the theory's assertion that only communities with sufficient demographic mass can sustain replacement levels and attract in-migration. On the other hand, ODR had positive correlates with housing prices, suggesting increased demand for housing among the elderly population. However, this effect diminished as ODR reached a threshold of 0.36, suggesting that once a municipality's population is overwhelmingly elderly, limited purchasing power and reduced household turnover begin to dampen market activity. These results suggest that once demographic thresholds are crossed, whether in density or age structure, municipalities may enter reinforcing decline trajectories. This echoes findings from trajectory-based studies that identify persistent demographic aging and weak population mass as structural drivers of long-term stagnation and aligns with research on institutional lock-ins (Aurambout et al., 2021; Turok and Mykhnenko, 2007) which in turn limit the capacity of shrinking municipalities to adapt (Grundel and Magnusson, 2023). This dynamic also reflects the central feedback mechanisms in CCT where demographic disadvantages amplify over time (Haase et al., 2014; Myrdal, 1960). Overall, the combined feature-importance of demographic factors is over 29% in the HPC model, reinforcing the idea that demographics are key determinants of housing market dynamics.

These findings reveal how demographic, economic, and housing-related mechanisms interact in self-reinforcing ways that deepen municipal shrinkage. Threshold effects, spatial disparities, and life-cycle dynamics jointly contribute to urban trajectories. As such, understanding shrinkage through the lens of systemic feedback loops offers a clearer view of why certain areas struggle despite national growth.

6.2 Methodological Insights

As the results section illustrated, when predicting shrinkage effects in Swedish municipalities, the Random Forest models achieved substantially higher model performance and lower error for PCR than for HPC. In part, this likely reflects the more stable patterns observed in PCR distributions (Section 5.1): between 2016–2020 and 2019–2023, most municipalities shifted from growth to shrinkage in a predictable manner, with large urban centers maintaining resilience while smaller and rural areas increasingly declined. These trends are well captured by the chosen predictors, allowing PCR models to explain over 65% of variance with relatively low RMSE. By contrast, HPC outcomes exhibited far greater dispersion, particularly in 2016–2020 before settling into more compressed, moderate house-price-growth in 2019–2023 (Fig. 2). This compression likely arose because national fiscal and monetary policy responses to COVID affected all municipalities simultaneously, causing housing markets to move in a more synchronized manner and thereby reducing cross-municipal variation in HPC. Consequently, the HPC model's RMSE is larger and more variable across subsets, reflecting that housing prices are driven by volatile, short-term shocks—factors a cross-sectional framework struggles to capture (Shmueli, 2010)—whereas population trajectories follow more stable medium-term patterns.

Analysing the variables using feature-importance and PDPs together offers a more complete picture than either approach alone. For instance, although PD received a low overall importance score in, its PDP highlights a clear threshold effect: PCR rises sharply below a PD thresholds of 100 inhabitants/km² and then flattens out, suggesting that PD is predictive mainly in sparsely populated areas, with minimal additional signal once that threshold is passed. Likewise, TP shows only modest feature-importance compared to variables like HSC_lag3 or RTP. Its PDP adds context to that by illustrating it‘s modest positive correlation with PCR.

These two analyses also shed light on changes in model performance. When explanatory power declines, as seen in the HPC model performance in the second period (Section 5.2.1), importance scores tend to flatten, as feature-importance scores show for the HPC predictors (Table 10), indicating that no single predictor captures much predictive signal. By contrast, a well-performing model like PCR in the first period (Table 9) shows a clear separation between high- and low-importance variables. This pattern carries over to the PDPs for HPC: in 2016–2020, predictors such as INC, HPC_lag3, and ODR exhibit sharp inflection points corresponding to known market mechanics, whereas in 2019–2023 those same curves become much flatter and noisier, reflecting diminished explanatory power.

Together, these insights demonstrate how combining feature-importance rankings with PDPs not only identifies the most influential predictors overall but also pinpoints the specific value ranges where those predictors drive model performance. In practice, this dual approach highlight where additional data or modelling strategies may be needed to understand evolving market conditions.

6.3 Performance Limitations

A key reason for the HPC model's weaker performance is the fragmented and spatially uneven nature of housing data on the municipal-level. Unlike demographic and economic indicators from SCB, housing data is scattered across sources with differing definitions, coverage, and formats, limiting comparability, while granularity varies sharply by geography. Metropolitan areas like Stockholm and Gothenburg provide disaggregated data due to frequent transactions, while rural municipalities often lack sufficient activity to generate meaningful metrics. This imbalance can bias the model toward urban patterns, since the RF algorithms learn from the data distributions they are exposed to, as noted by (Breiman, 2001). An overrepresentation of urban areas may hence lead to model structures that generalise poorly in rural contexts, while masking structural vacancy or fragility that are more prevalent but less visible in low-data regions.

As a result, the analysis in this study relies on proxies like RTP and DSC_lag3 which offer only coarse approximations of local dynamics. These variables may not fully capture aspects such as renovation activity, speculative transactions, short-term letting volumes, or the role of municipally owned housing.

Temporal constraints further complicate model performance. Housing variables used in this study are reported annually and with delay, limiting the model's responsiveness to short-term shocks or market swings. Combined with the cross-sectional design, it reduces the model's ability to capture time-lagged or cyclical effects (Shmueli, 2010), especially in volatile housing markets. As Breiman (2001) emphasises, predictive algorithms like RF rely on high-quality, granular data – coarse or uneven inputs hinder its ability to uncover consistent patterns. Data of this quality would enhance sensitivity to local trends and enable sub-municipal housing market analysis. Improved data access and variable availability would likely boost model accuracy (Shmueli, 2010) for a cross-sectional analysis of HPC like the one employed in this study.

At the same time, the consistent predictive accuracy of the PCR model across periods affirms the value of machine learning approaches in capturing structural demographic trends with a cross-sectional design. While this design enabled broad structural comparisons, future work could benefit from a longitudinal approach to capture within-municipality dynamics, account for temporal lags, and better isolate causal feedback effects that unfold over time.

6.4 Policy and Research Implications

The identification of clear thresholds and nonlinear dynamics in both PCR and HPC models carries important implications for policy. First, the strong predictive role of RTP and DSC_lag3 suggests a need of affordability and supply-side interventions in vulnerable municipalities. National housing strategies should prioritize rental subsidies or cost-containment measures in localities where rent burdens cross tipping points where out-migration accelerates, eroding municipal tax bases. At the same time, targeted support for moderate-scale construction, enough to exceed the construction thresholds but avoid oversupply, could help stabilize demographic decline by signalling market confidence without triggering downward pressure on prices. Second, the threshold effects of scale (TP < 50,000; PD < 100 inhabitants/km²) suggest that smaller sparsely settled municipalities may require differentiated governance tools. Rather than one-size-fits-all planning, local governments in these areas could benefit from proactive measures to manage shrinkage, rather than pursuing growth – a perspective also highlighted by Grundel and Magnusson (2023).

7 Conclusion

This study set out to (1) evaluate how accurately cross-sectional Random Forest regression models can predict municipal population change (PCR) and housing-price change (HPC) in Sweden, using demographic, economic and housing indicators, and (2) to uncover the nonlinear relationships between predictor variables and outcome variables and the structural, temporal complexities of urban shrinkage in Swedish municipalities.

7.1 Summary of key findings

The RF models showed greater predictive performance for PCR than for HPC across both study periods. The PCR models performed best on the full sample, achieving R^2 values of 0.67 in 2016–2020 and 0.65 in 2019–2023, with corresponding RMSE values stable around 0.02 in both periods. In contrast, the HPC model demonstrated substantially lower explanatory power. During the second period, it failed to produce meaningful results across all samples; however, when municipalities with fewer than 10,000 inhabitants were excluded, the model exhibited modest predictive ability (R^2 = 0.33; RMSE = 0.085). This performance gap suggests that structural demographic, economic, and housing indicators are more effective in explaining long-term population trends than in capturing housing price dynamics. The latter may be influenced by exogenous shocks and spatial heterogeneities, which are difficult to capture within the limitations of a cross-sectional Random Forest framework.

Theoretical analyses of the feature-importance scores and partial dependence plots (PDPs) of the variables contributed to deepening the understanding of the structural, temporal complexities of urban shrinkage in Sweden. For the PCR model, two of the three top-ranked variables are housing-related, whereas in the HPC model, two out of the four leading variables are demographic. When interpreting the results, it reveals clear interconnections between the housing and demographic domains, thereby reinforcing the theoretical perspective of feedback loops across these systems. Among economic indicators, only average disposable income (INC) exerts a notable influence. In comparison, employment rate (ER) and Gross Domestic Product per capita (GDP_pc) were less influential than anticipated.

Across both periods, the PDPs from the PCR model reveal that municipal shrinkage does not progress in a smooth, linear fashion but instead hinges on clear threshold effects, critical values at which demographic, economic, or housing-market conditions sharply shift from stability to decline. Interpreting these threshold effects through the theoretical framework contributed to understanding the complexities of urban shrinkage in Sweden. For instance, as discussed in Section 6.1, population outflow appears to be triggered when RTP exceeds a ratio of 0.35, and a similar pattern was observed when ODR surpassed a ratio of 0.35 as well. Overall, the nonlinear relationships highlighted in Section 5.1 indicate that urban shrinkage is triggered when structural pressures, such as aging populations, inadequate housing, converge past manageable tipping points.

The PDPs of the HPC model exhibit more distinct and interpretable threshold effects during the first period. In contrast, in the 2019-2023 period, the same variables produced flatter and noisier curves, offering limited predictive insight. Notably, prior house-price changes (HPC_lag3) in the 2016–2020 period, together with municipal income levels (INC), accounted for roughly half of the HPC model's explanatory power—showing that price momentum and income thresholds drive nonlinear shifts from growth to maturity in local markets, which resonates with the study's theoretical framework. These patterns suggest that once local incomes surpassed certain thresholds or previous price accelerations peaked, municipalities tended to shift rapidly from booming to stagnating housing markets. However, in the 2019–2023 period, the predictive power of structural indicators weakened considerably. This likely reflects heightened volatility and systemic shifts following the COVID pandemic, which affected all municipalities simultaneously, causing housing markets to move in more similar fashion, leaving less differentiating signal for the RF model to exploit. This contrast underscores a key finding: while demographic thresholds consistently serve as reliable indicators of population decline, the predictability of housing-price dynamics appears more vulnerable to disruption during periods of macroeconomic turbulence. More broadly, these threshold effects highlight the value of tracking a few critical "red-flag" indicators, as crossing certain bounds can trigger feedback mechanisms that accelerate shrinkage and pose challenges for gradual policy responses.

7.2 Future Research

This study contributes to the growing field of machine learning in urban analytics by integrating predictive modelling with theory-informed interpretation, demonstrating the potential of Random Forests to capture nonlinear dynamics associated with structural demographic change. At the same time, the limitations of cross-sectional data—particularly in modelling volatile outcomes like housing prices—highlight the need for future research that leverages longitudinal designs, finer temporal granularity, and spatially richer datasets. The unexplained variance in housing price change suggests that important drivers remain unaccounted for, including spatial spillovers, speculative investment behaviour, and land-use regulation. Incorporating proxies for policy interventions (e.g., zoning changes, infrastructure investment) and household-level factors (e.g., sentiment, resilience) could help capture market frictions and institutional responses. Additionally, expanding the variable set beyond publicly available indicators, such as through mixed-methods or qualitative integration, may offer critical insights into governance capacity and local adaptation. By combining machine learning with enriched spatial, temporal, and institutional data, future research can more effectively support anticipatory planning and context-sensitive strategies for managing urban decline.

8 Reference List

Aurambout, J.P., Schiavina, M., Melchiori, M., Fioretti, C., Guzzo, F., Vandecasteele, I., Proietti, P., Kavalov, B., Panella, F., Koukoufikis, G., 2021. Shrinking Cities (No. JRC126011), The Future of Cities. European Commission, Joint Research Centre (JRC), Luxembourg.

Beauregard, R.A., 2009. Urban Population Loss in Historical Perspective: United States, 1820–2000. Environ Plan A 41, 514–528. https://doi.org/10.1068/a40139a

Bloom, D.E., Canning, D., Fink, G., 2010. Implications of population ageing for economic growth. Oxford Review of Economic Policy 26, 583–612. https://doi.org/10.1093/oxrep/grq038

Boverket, 2024. Bostadsmarknadsenkäten - Riket [WWW Document]. Boverket. URL https://www.boverket.se/sv/samhallsplanering/bostadsmarknad/bostadsmarknaden/bostadsmarknadsenkaten/region-kommun/riket/ (accessed 4.26.25).

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Carson, D.B., Carson, D.A., Porter, R., Ahlin, C.Y., Sköld, P., 2016. Decline, adaptation or transformation: New perspectives on demographic change in resource peripheries in Australia and Sweden. Comparative Population Studies 41. https://doi.org/10.12765/CPoS-2016-11en

Doeringer, S., Uchiyama, Y., Penker, M., Kohsaka, R., 2020. A meta-analysis of shrinking cities in Europe and Japan. Towards an integrative research agenda. Eur. Plan. Stud. 28, 1693–1712. https://doi.org/10.1080/09654313.2019.1604635

Eimermann, M., Adjei, E.K., Bjarnason, T., Lundmark, L., 2022. Exploring population redistribution at sub-municipal levels – Microurbanisation and messy migration in Sweden's high North. Journal of Rural Studies 90, 93–103. https://doi.org/10.1016/j.jrurstud.2022.01.010

Grundel, I., Magnusson, D., 2023. Planning to grow, planning to rock on – infrastructure management and development in shrinking municipalities. European Planning Studies 31, 1184–1202. https://doi.org/10.1080/09654313.2022.2108311

Haase, A., Bernt, M., Großmann, K., Mykhnenko, V., Rink, D., 2016. Varieties of shrinkage in European cities. European Urban and Regional Studies 23, 86–102. https://doi.org/10.1177/0969776413481985

Haase, A., Rink, D., Grossmann, K., Bernt, M., Mykhnenko, V., 2014. Conceptualizing urban shrinkage. Environment and Planning A 46, 1519–1534. https://doi.org/10.1068/a46269

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, Springer Series in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-0-387-84858-7

He, S.Y., Lee, J., Zhou, T., Wu, D., 2017. Shrinking cities and resource-based economy: The economic restructuring in China's mining cities. Cities 60, 75–83. https://doi.org/10.1016/j.cities.2016.07.009

Hedlund, M., Lundholm, E., 2015. Restructuring of rural Sweden - Employment transition and out-migration of three cohorts born 1945-1980. Journal of Rural Studies 42, 123–132. https://doi.org/10.1016/j.jrurstud.2015.10.006

Kato, H., 2024. Multidimensional factors correlated with population changes according to city size in Japan. Env. Plan. B-Urban Anal. City Sci. https://doi.org/10.1177/23998083241274381

Khavarian-Garmsir, A.R., 2023. A systematic review of shrinking cities literature: lessons from the past and directions for the future. International Planning Studies 28, 219–238. https://doi.org/10.1080/13563475.2023.2205030

Kirk, D., 1996. Demographic Transition Theory. Population Studies 50, 361–387. https://doi.org/10.1080/0032472031000149536

Kitchin, R., 2014. Big Data, new epistemologies and paradigm shifts. Big Data & Society 1, 2053951714528481. https://doi.org/10.1177/2053951714528481

Lee, R., 2003. The demographic transition: Three centuries of fundamental change. J. Econ. Perspect. 17, 167–190. https://doi.org/10.1257/089533003772034943

Lindh, T., Malmberg, B., 2008. Demography and Housing Demand—What Can We Learn from Residential Construction Data? Journal of Population Economics 21, 521–539.

Liu, Z., Liu, S., Song, Y., 2020. Understanding urban shrinkage in China: Developing a multi-dimensional conceptual model and conducting empirical examination from 2000 to 2010. Habitat International 104, 102256. https://doi.org/10.1016/j.habitatint.2020.102256

Lundmark, L., Carson, D.A., Eimermann, M., 2022. Spillover, sponge or something else? Dismantling expectations for rural development resulting from giga-investments in Northern Sweden. Fennia 200, 157–174. https://doi.org/10.11143/fennia.120530

Ma, J., Cheng, J.C.P., Jiang, F., Chen, W., Zhang, J., 2020. Analyzing driving factors of land values in urban scale based on big data and non-linear machine learning techniques. Land Use Policy 94, 104537. https://doi.org/10.1016/j.landusepol.2020.104537

Ma, Z., Zhou, G., Zhang, J., Liu, Y., Zhang, P., Li, C., 2024. Urban shrinkage in the regional multiscale context: Spatial divergence and interaction. Sustainable Cities and Society 100, 105020. https://doi.org/10.1016/j.scs.2023.105020

Malmberg, B., 2010. Low Fertility and the Housing Market: Evidence from Swedish Regional Data / Basse Fécondité et Marché du Logement: Une Analyse de Données Régionales Suédoises. European Journal of Population / Revue Européenne de Démographie 26, 229–244.

Martinez-Fernandez, C., Audirac, I., Fol, S., Cunningham-Sabot, E., 2012. Shrinking Cities: Urban Challenges of Globalization. International Journal of Urban and Regional Research 36, 213–225. https://doi.org/10.1111/j.1468-2427.2011.01092.x

Myrdal, 1960. By Gunnar Myrdal. Bombay, Vora, 1958. pp. 183. Rs. 4.50. Quarterly 16, 189–191. https://doi.org/10.1177/097492846001600220

Nordregio, n.d. The spatial planning systems in the Nordic region - Nordregio [WWW Document]. URL https://archive.nordregio.se/Metameny/About-Nordregio/Nordic-working-groups/nwgcityregions/The-spatial-planning-systems-in-the-Nordic-region/index.html (accessed 5.21.25).

OECD, 2023. Government at a Glance 2023: Sweden. OECD Publishing, Paris.

OECD, 2019. Digital Government Review of Sweden: Towards a Data-driven Public Sector, OECD Digital Government Studies. OECD Publishing, Paris.

Peng, W., Wu, Z., Duan, J., Gao, W., Wang, R., Fan, Z., Liu, N., 2023. Identifying and quantizing the non-linear correlates of city shrinkage in Japan. Cities 137, 104292. https://doi.org/10.1016/j.cities.2023.104292

Reckien, D., Martinez-Fernandez, C., 2011. Why do cities shrink? European Planning Studies 19, 1375–1397. https://doi.org/10.1080/09654313.2011.593333

Regeringen, 2025. Ny strategi för levande och trygga städer (No. Skr. 2024/25:96). Regeringskansliet, Stockholm.

Saraiva, M., Roebeling, P., Sousa, S., Teotónio, C., Palla, A., Gnecco, I., 2017. Dimensions of shrinkage: Evaluating the socio-economic consequences of population decline in two medium-sized cities in Europe, using the SULD decision support tool. Environment and Planning B 44, 1122–1144. https://doi.org/10.1177/0265813516659071

Sassen, S., 2001. The global city: New York, London, Tokyo, 2. ed. ed. Princeton University Press, Princeton, N.J.

SCB, 2024a. Sveriges framtida befolkning 2024-2070 (No. 2024:1), Demografiska Rapporter. Statistiska centralbyrån (SCB), Stockholm.

SCB, 2024b. Lägsta folkökningen på 22 år, Befolkningsstatistik helåret 2023. Statistiska centralbyrån (SCB).

SCB, 2024c. Samordning av Sveriges officiella statistik [WWW Document]. Statistikmyndigheten SCB. URL https://www.scb.se/om-scb/samordning-av-sveriges-officiella-statistik/ (accessed 5.1.25).

Schilling, J., and Logan, J., 2008. Greening the Rust Belt: A Green Infrastructure Model for Right Sizing America's Shrinking Cities. Journal of the American Planning Association 74, 451–466. https://doi.org/10.1080/01944360802354956

Shmueli, G., 2010. To explain or to predict? Statistical Science 25, 289–310. https://doi.org/10.1214/10-STS330

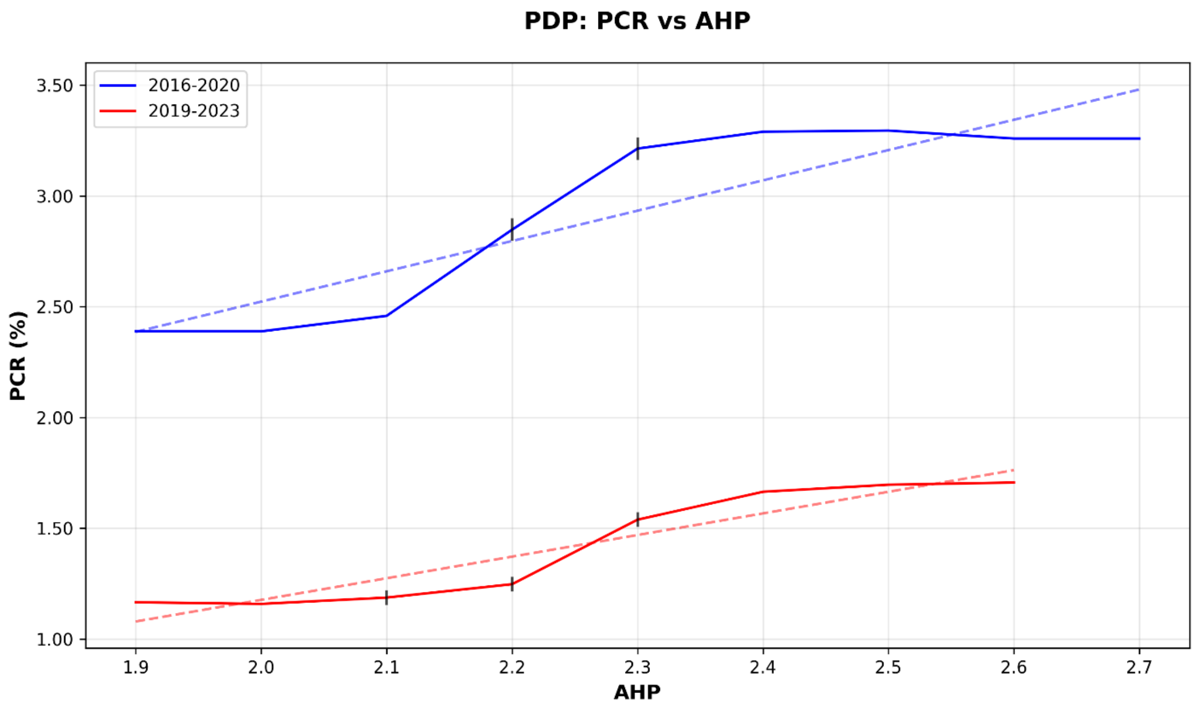
Tillväxtverket, 2024. Regionalt utvecklingsarbete 2023 (No. Rapport 0479). Tillväxtverket, Stockholm.

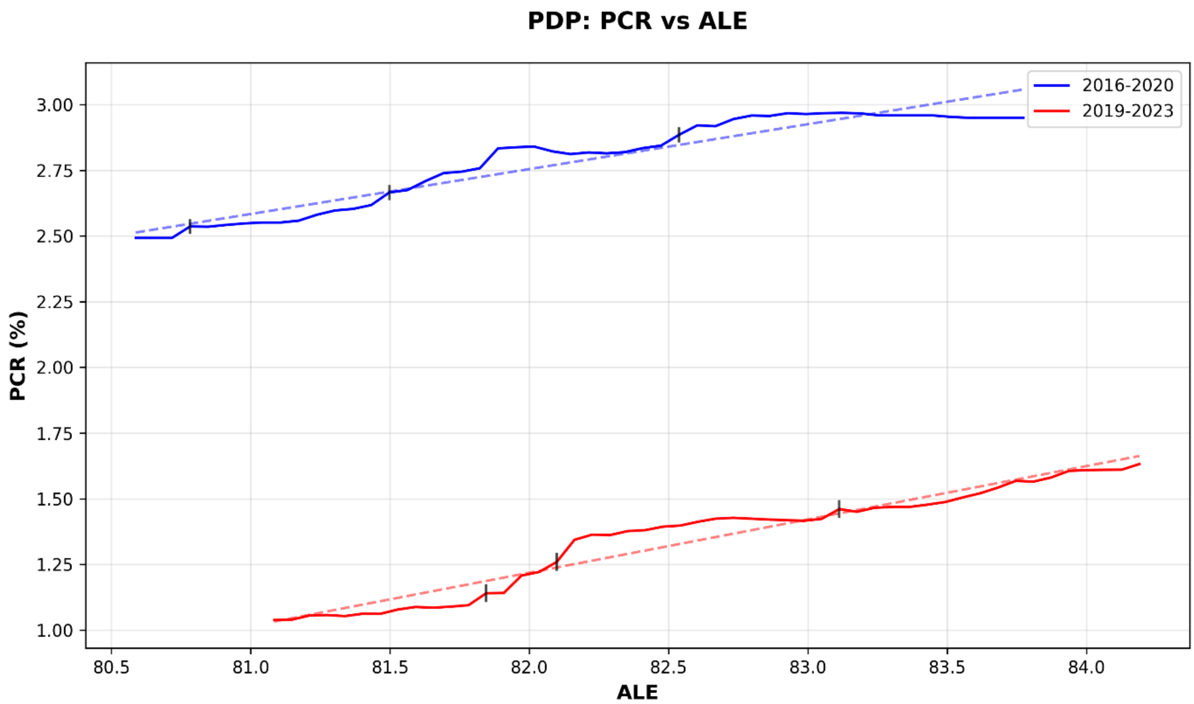
Turok, I., Mykhnenko, V., 2007. The trajectories of European cities, 1960–2005. Cities 24, 165–182. https://doi.org/10.1016/j.cities.2007.01.007

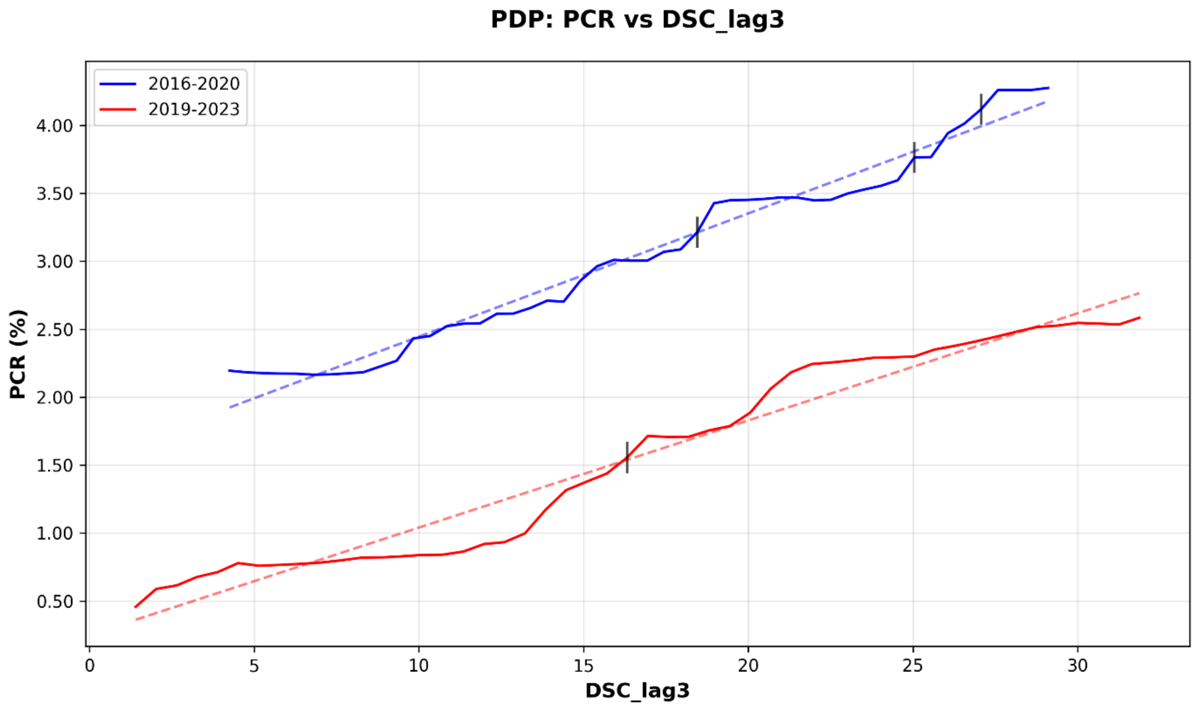
UN DESA, 2024. World Population Prospects 2024: Summary of Results (No. UN DESA/POP/2024/TR/NO. 9). United Nations, New York, NY.
Wang, J., Biljecki, F., 2022. Unsupervised machine learning in urban studies: A systematic review of applications. Cities 129, 103925. https://doi.org/10.1016/j.cities.2022.103925
Wang, L., Zhang, R., Liu, Z., Shirakawa, H., Tanikawa, H., 2024. From expansion to efficiency: Machine learning-based forecasting of Japan's building material stocks under demographic declines. Science of The Total Environment 951, 175634. https://doi.org/10.1016/j.scitotenv.2024.175634
Wiechmann, T., 2008. Errors Expected — Aligning Urban Strategy with Demographic Uncertainty in Shrinking Cities†. International Planning Studies 13, 431–446. https://doi.org/10.1080/13563470802519097
Wiechmann, T., Pallagst, K.M., 2012. Urban shrinkage in Germany and the USA: A Comparison of Transformation Patterns and Local Strategies. International Journal of Urban and Regional Research 36, 261–280. https://doi.org/10.1111/j.1468-2427.2011.01095.x
Zheng, Y., 2023. Community resilience and house prices: A machine learning approach. Finance Research Letters 58, 104400. https://doi.org/10.1016/j.frl.2023.104400

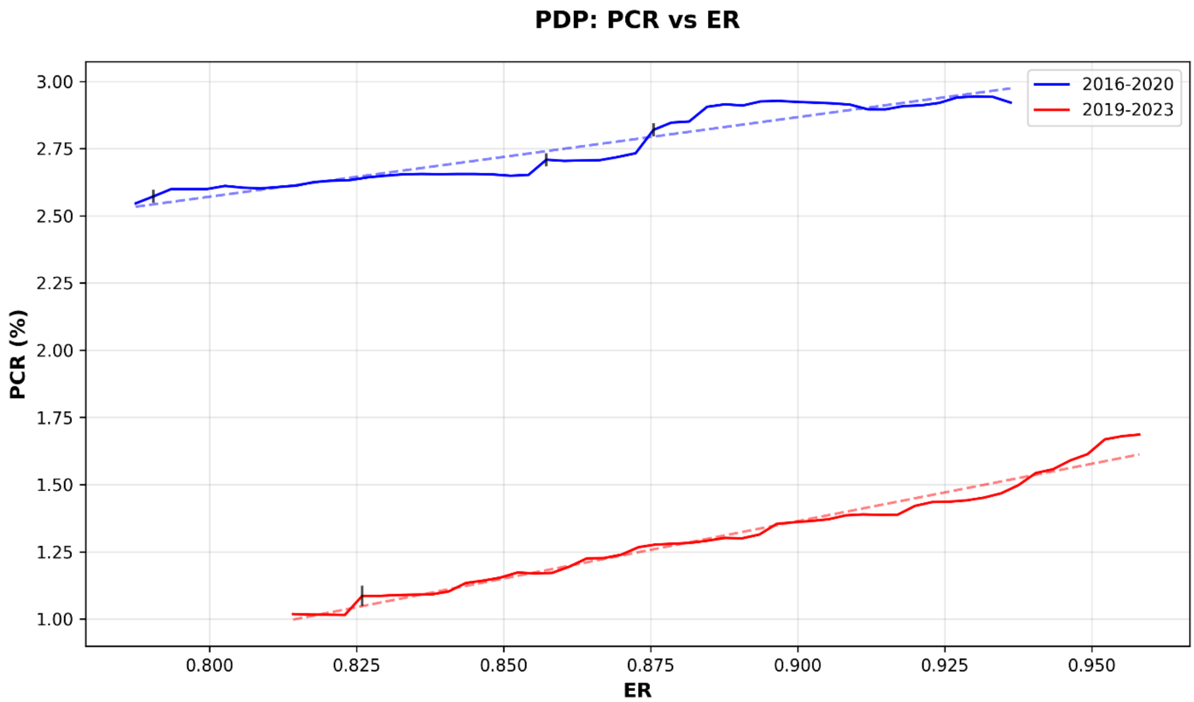
9 Appendix

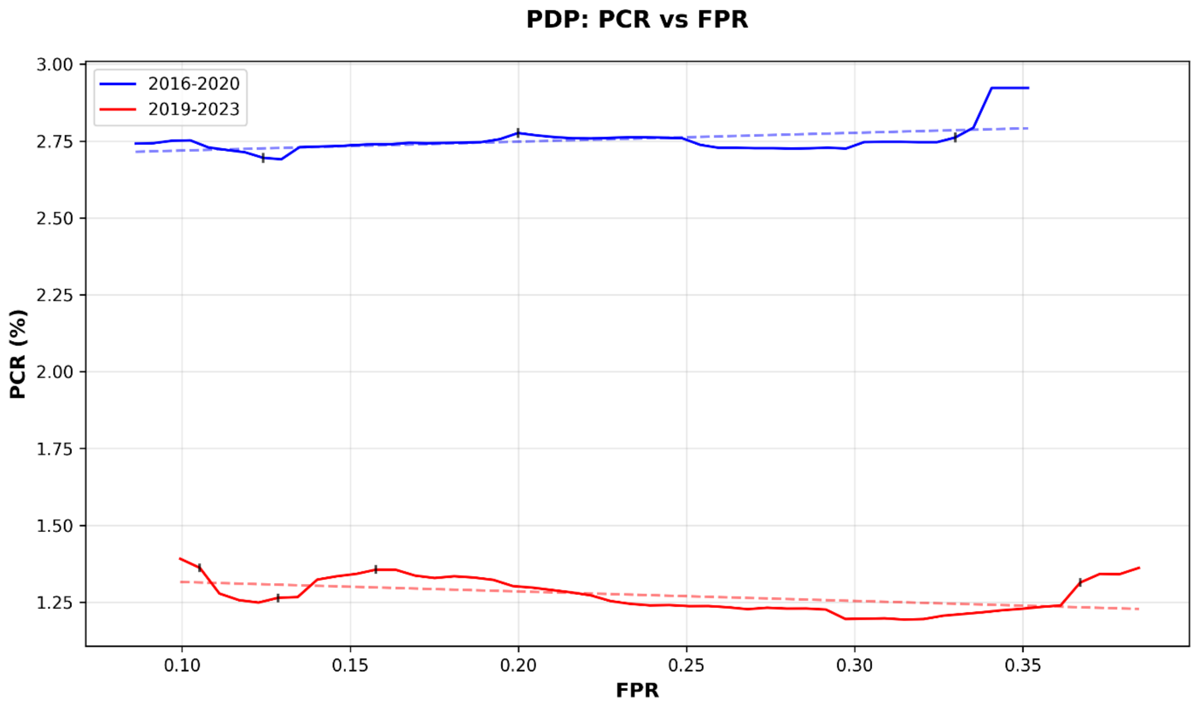
9.1 Appendix A: PCR Partial Dependence Plots



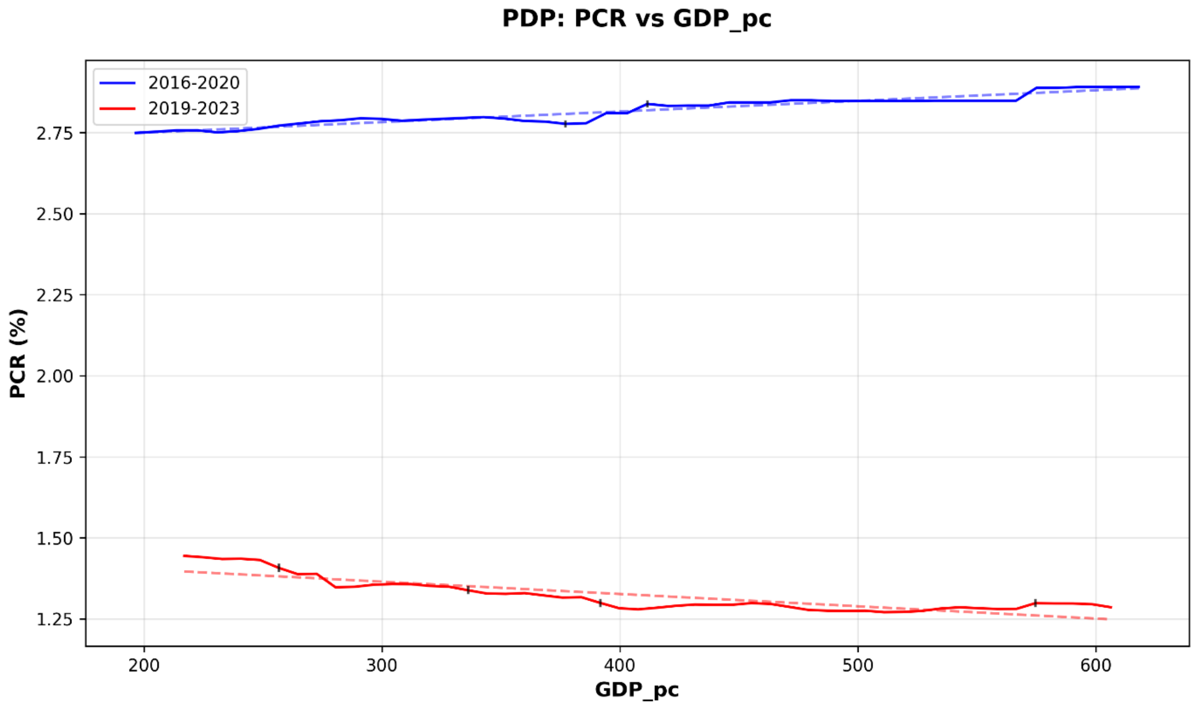



PDP: PCR vs DSC_lag3
2016-2020
2019-2023
PCR (%)
DSC_lag3
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0
5
10
15
20
25
30

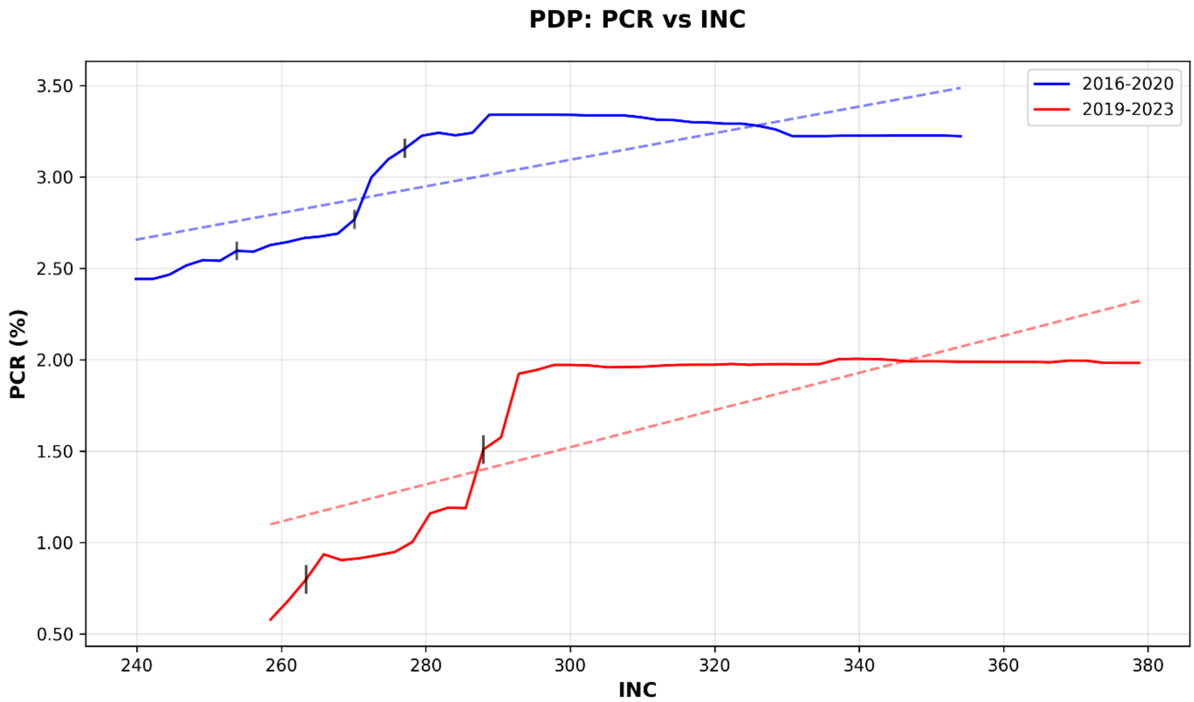

PDP: PCR vs ER
2016-2020
2019-2023
PCR (%)
ER
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.800
0.825
0.850
0.875
0.900
0.925
0.950

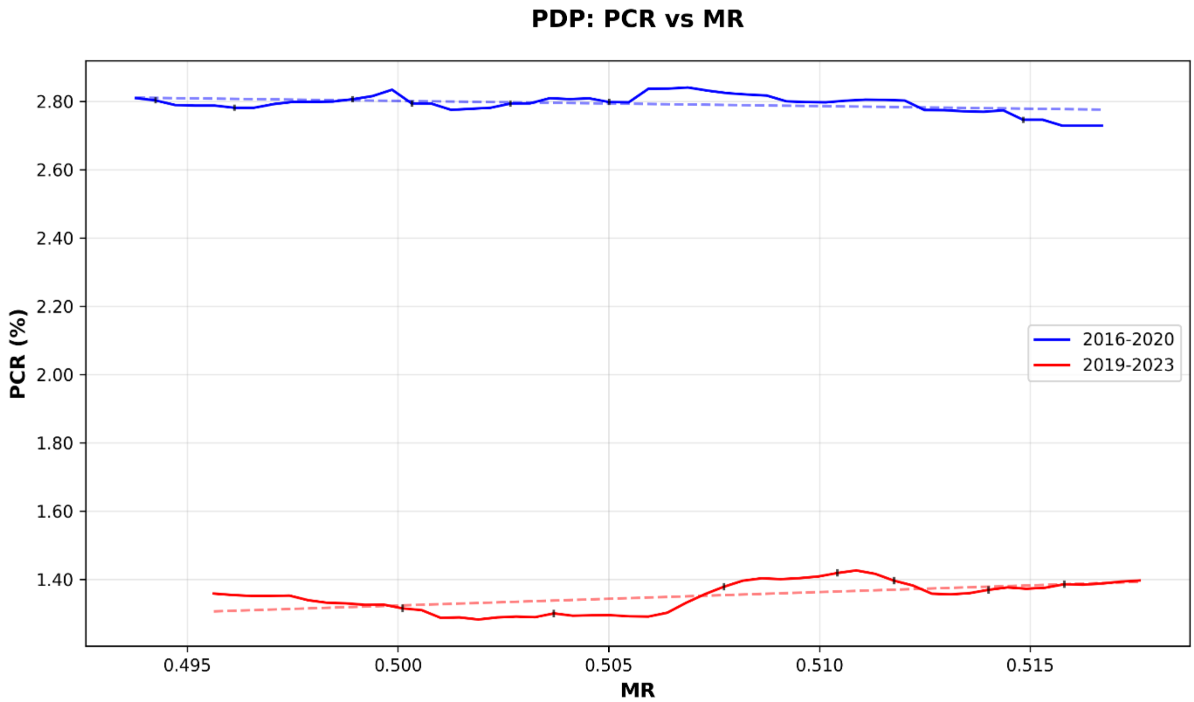


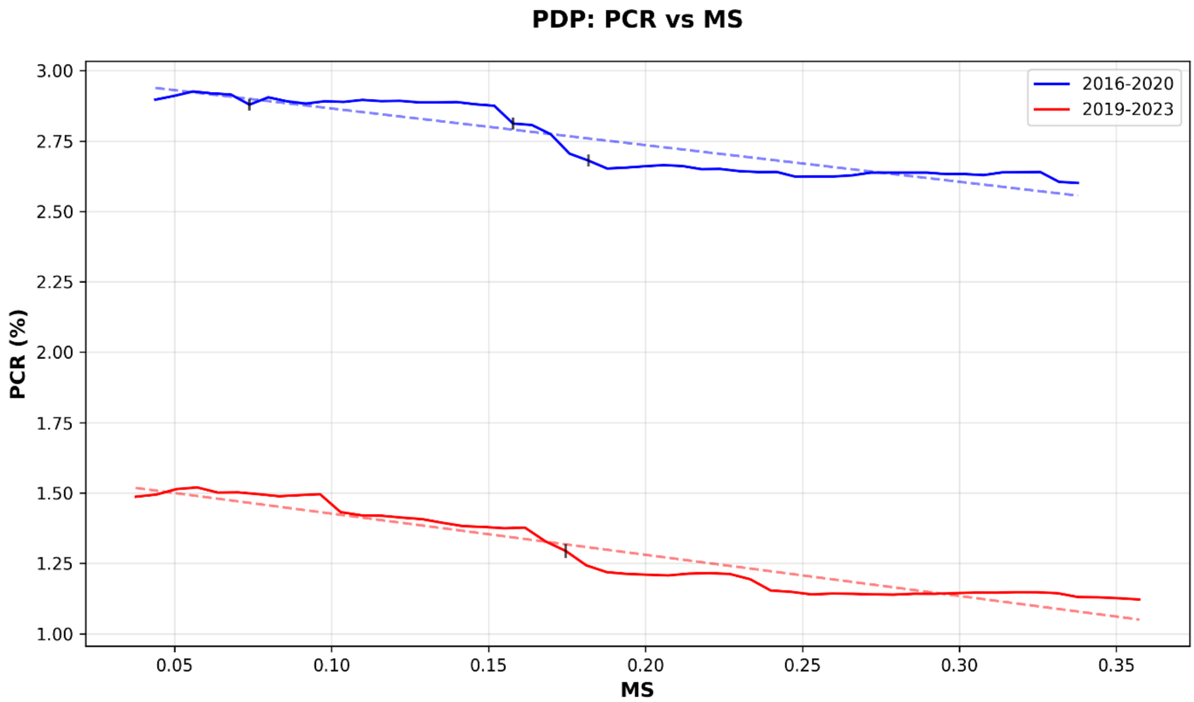
PDP: PCR vs FPR
2016-2020
2019-2023
PCR (%)
FPR

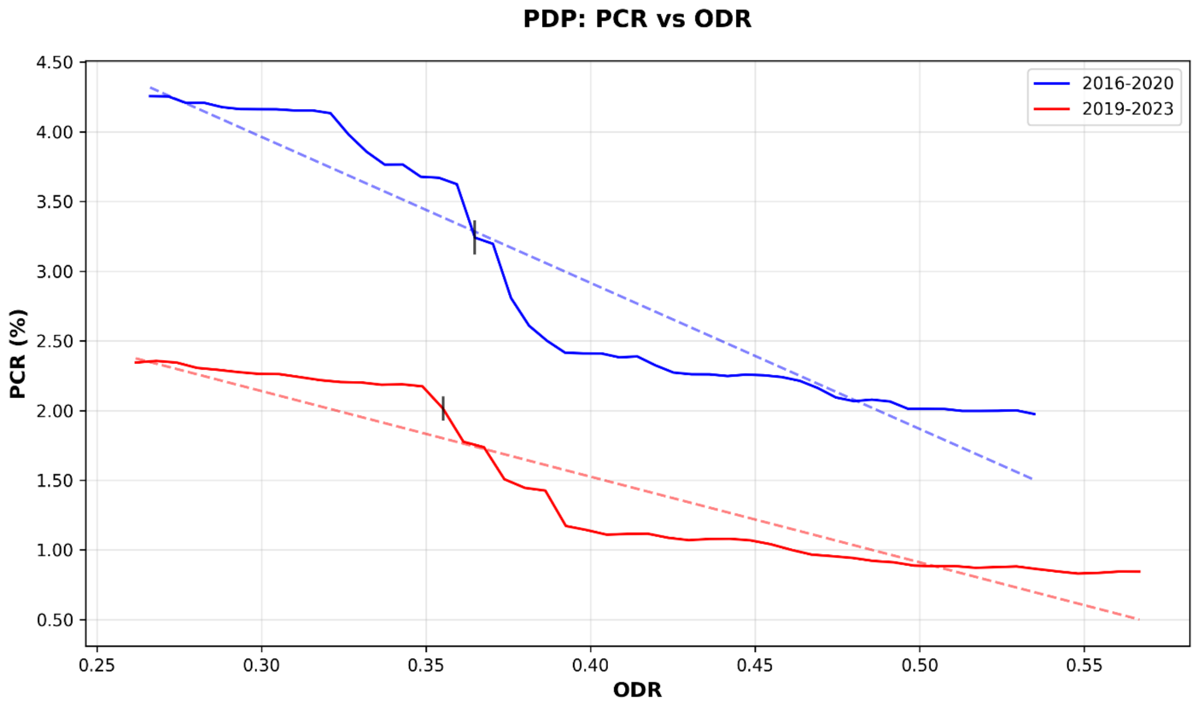


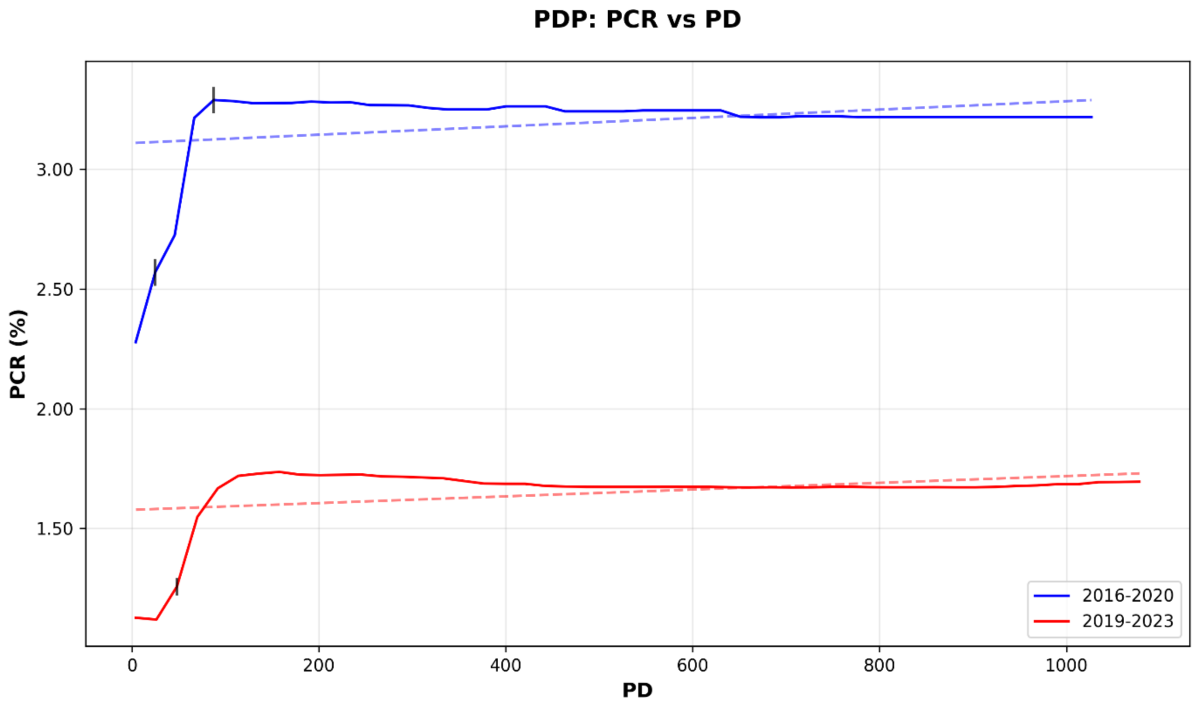
PDP: PCR vs GDP_pc
2016-2020
2019-2023
PCR (%)
GDP_pc

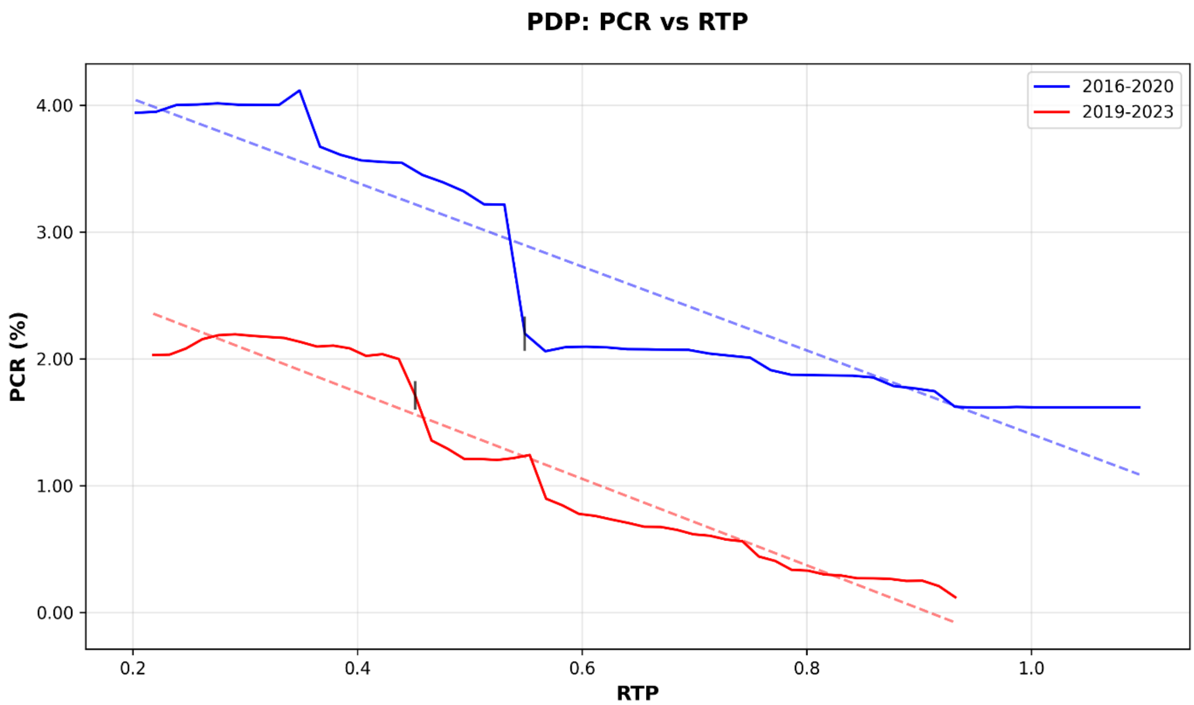

PDP: PCR vs INC
PCR (%)
INC
2016-2020
2019-2023
3.50
3.00
2.50
2.00
1.50
1.00
0.50
240
260
280
300
320
340
360
380

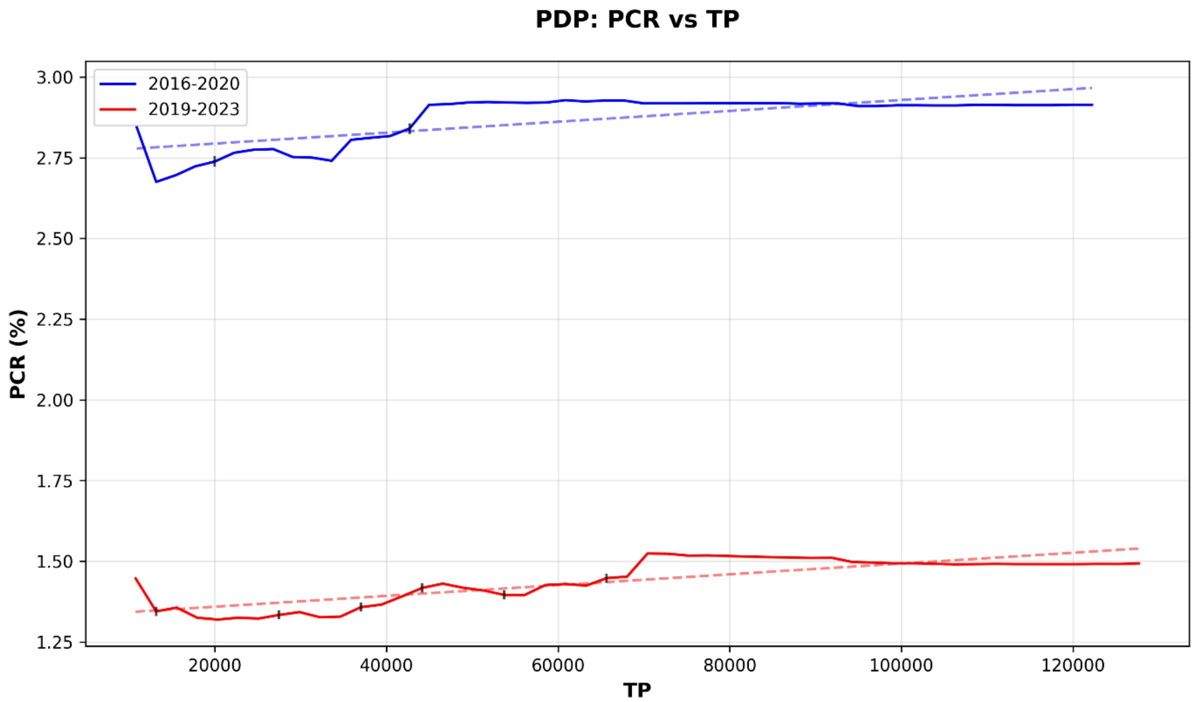

PDP: PCR vs MR
PCR (%)
MR
2016-2020
2019-2023
2.80
2.60
2.40
2.20
2.00
1.80
1.60
1.40
0.495
0.500
0.505
0.510
0.515

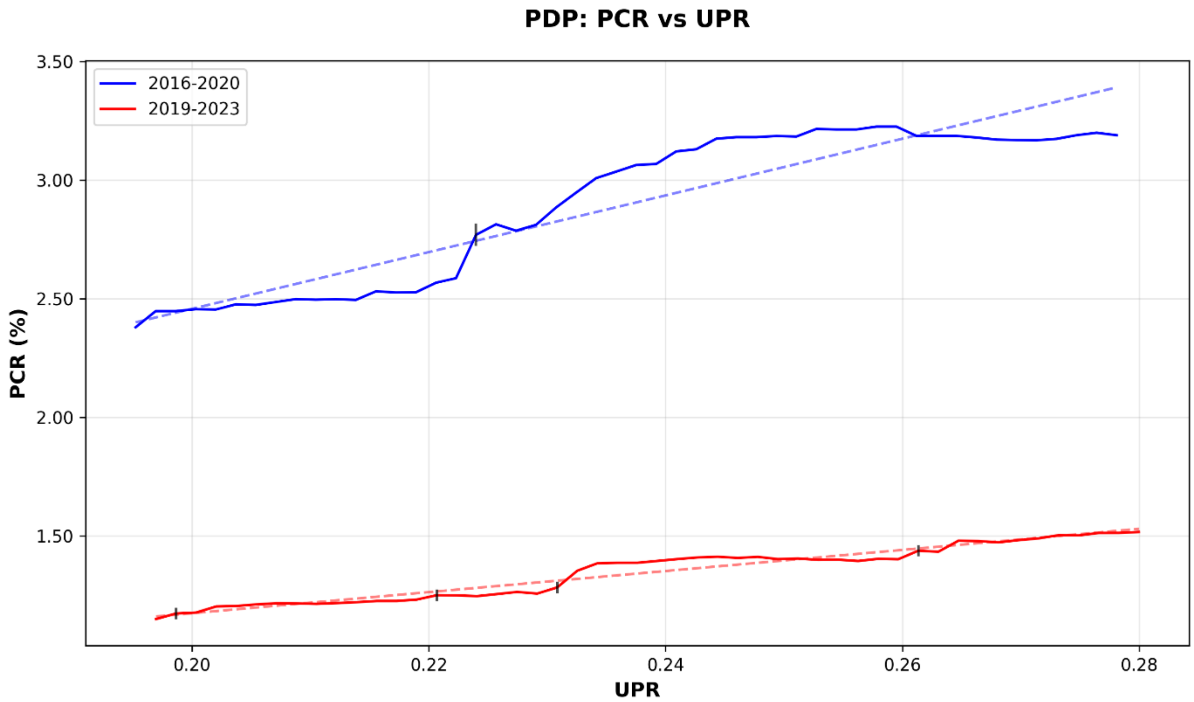

PDP: PCR vs MS
PCR (%)
MS
2016-2020
2019-2023
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35


PDP: PCR vs ODR
PCR (%)
ODR
2016-2020
2019-2023
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.25
0.30
0.35
0.40
0.45
0.50
0.55

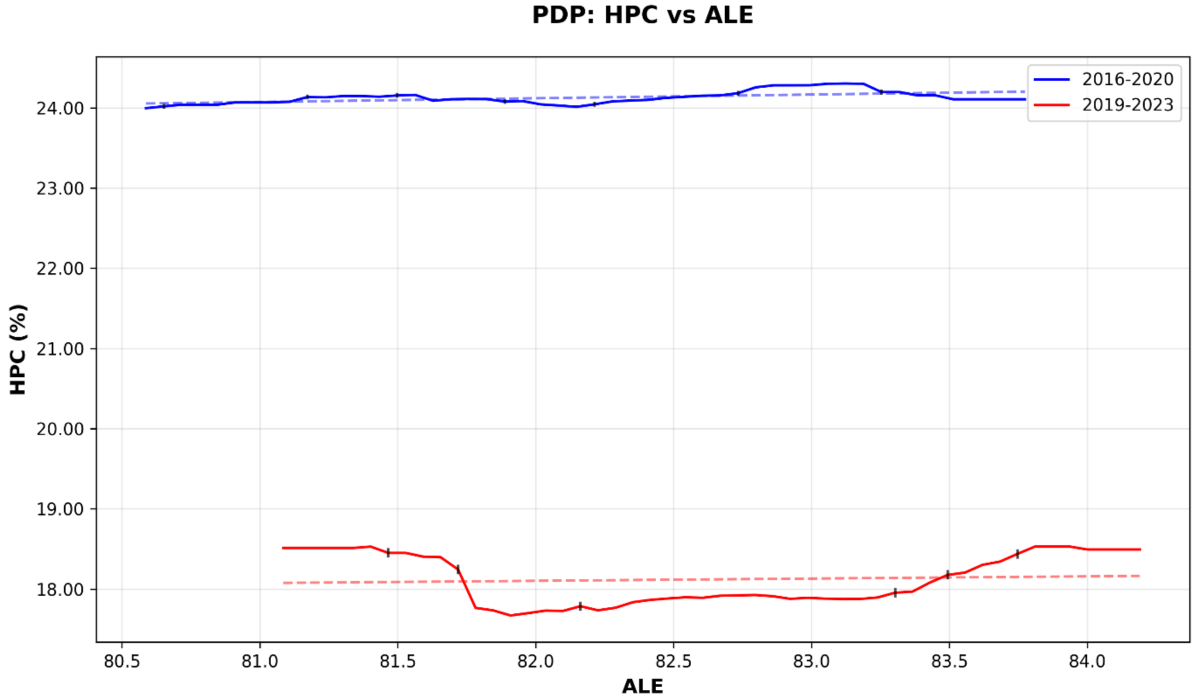

PDP: PCR vs PD
PCR (%)
3.00
2.50
2.00
1.50
0
200
400
600
800
1000
PD
2016-2020
2019-2023

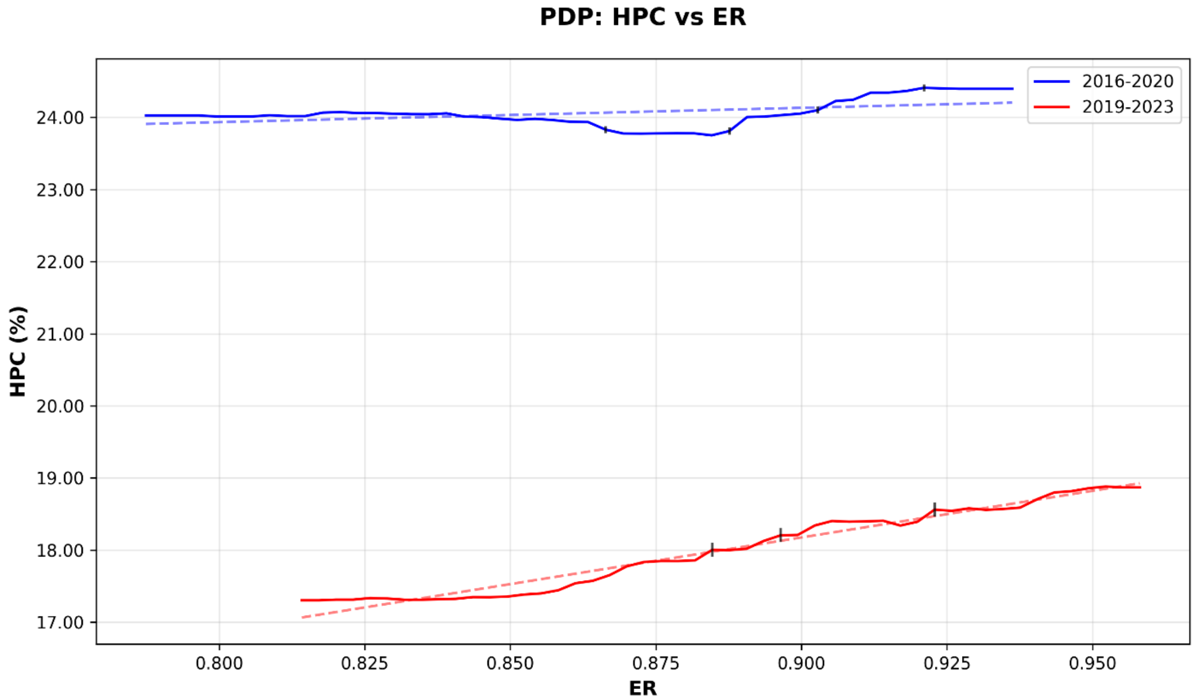

PDP: PCR vs RTP
PCR (%)
4.00
3.00
2.00
1.00
0.00
0.2
0.4
0.6
0.8
1.0
RTP
2016-2020
2019-2023

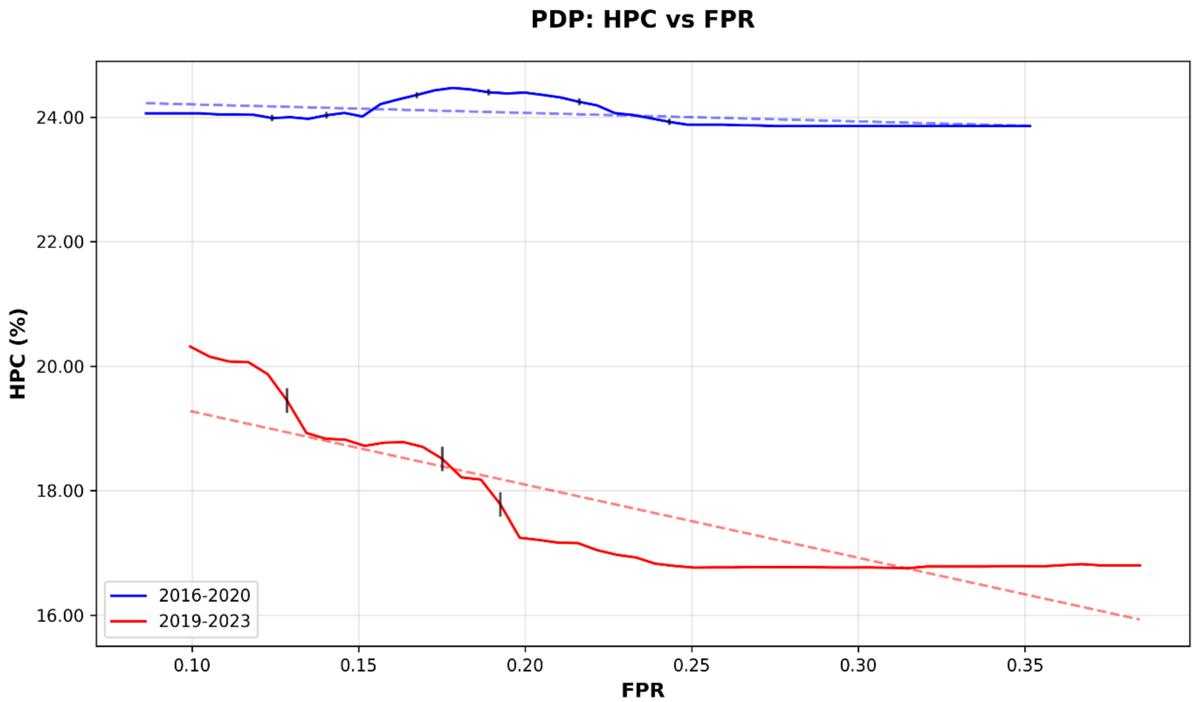

PDP: PCR vs TP
2016-2020
2019-2023
PCR (%)
TP
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
20000
40000
60000
80000
100000
120000

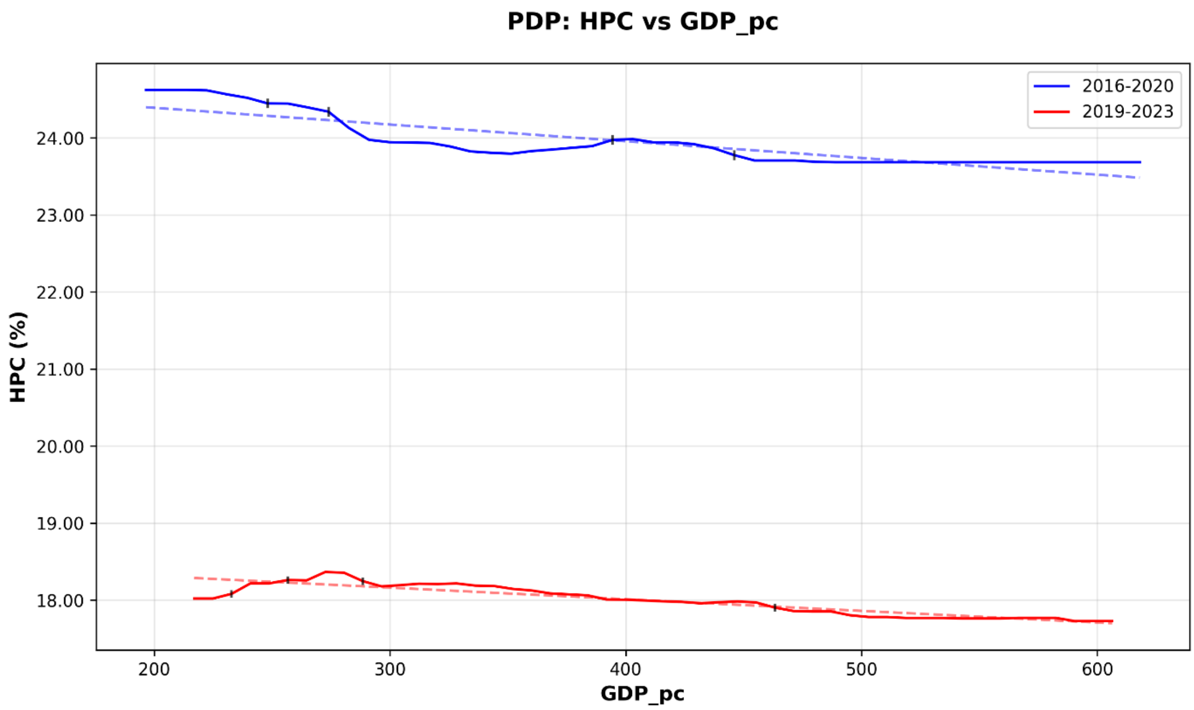

PDP: PCR vs UPR
2016-2020
2019-2023
PCR (%)
UPR
3.50
3.00
2.50
2.00
1.50
0.20
0.22
0.24
0.26
0.28

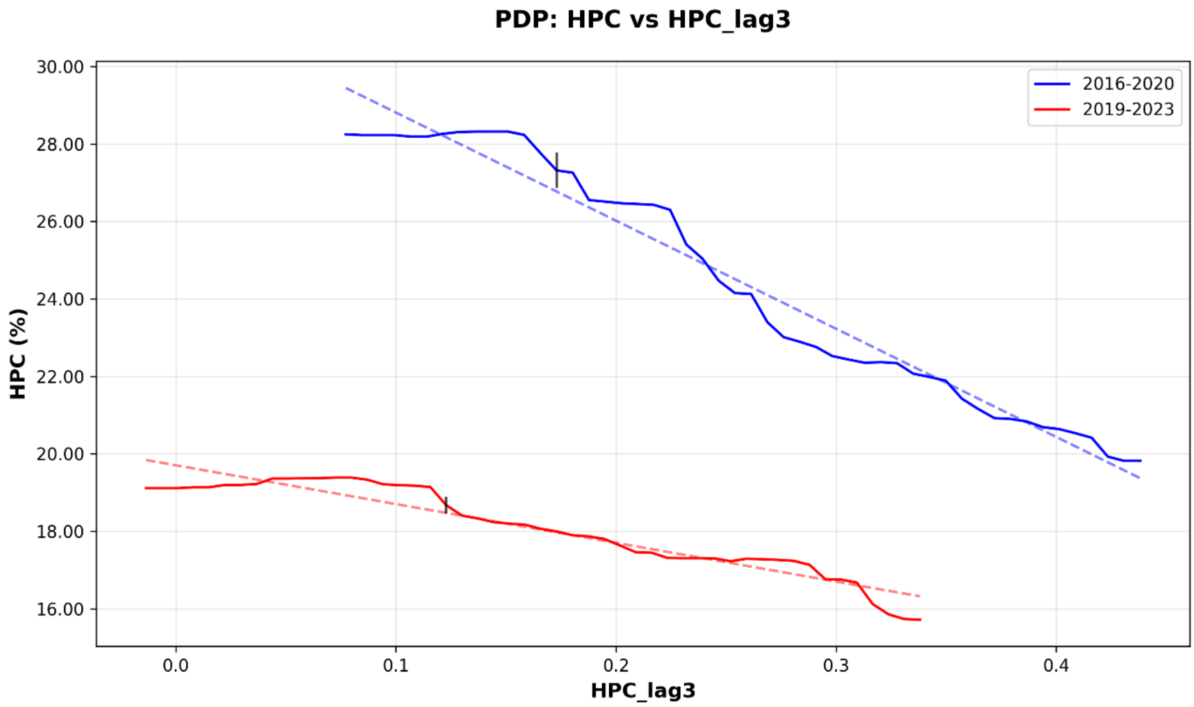
9.2 Appendix B: HPC Partial Dependence Plots

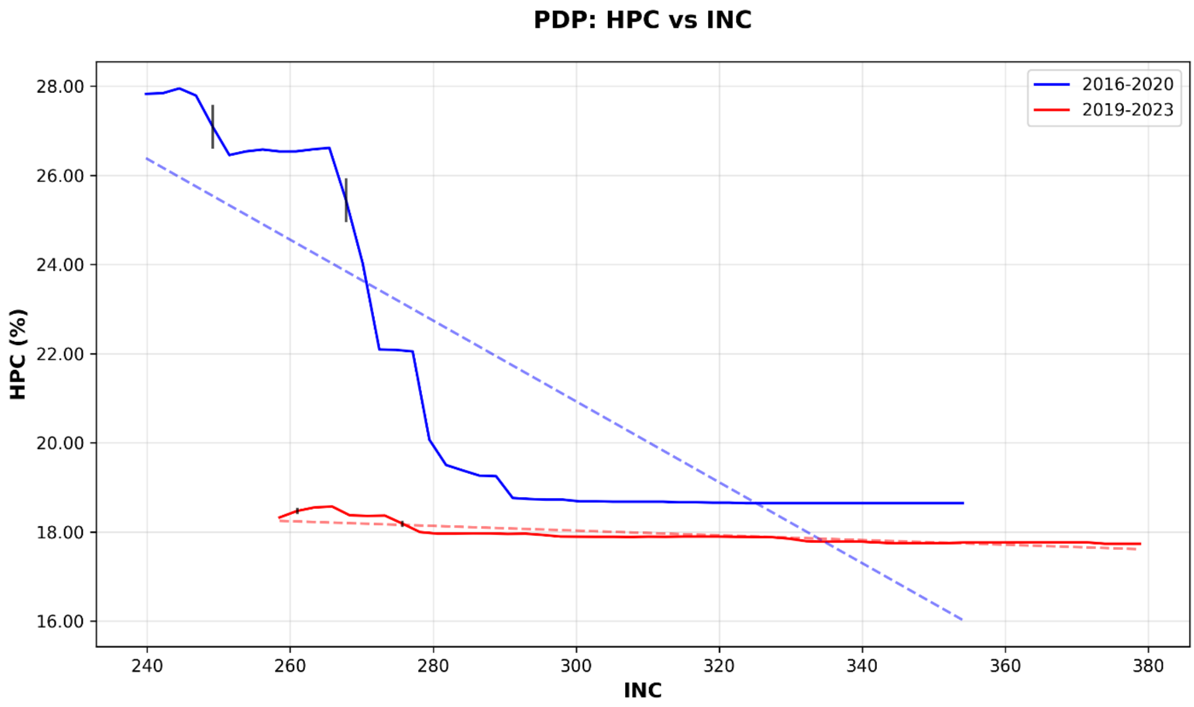


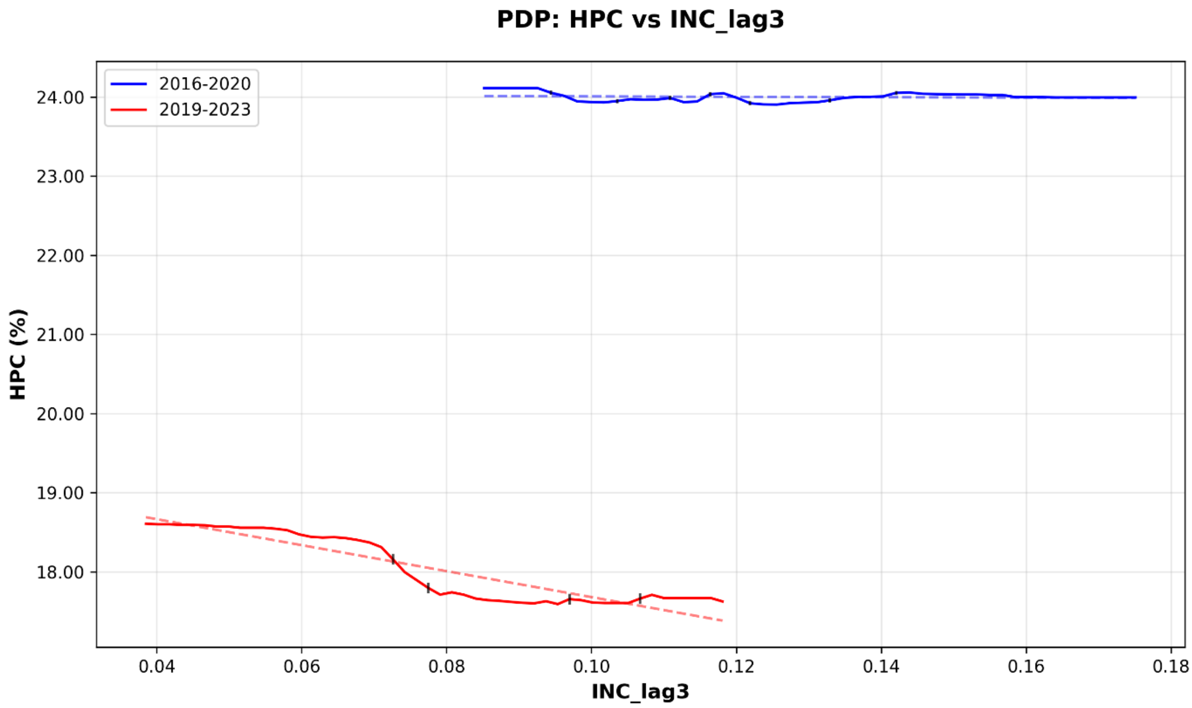


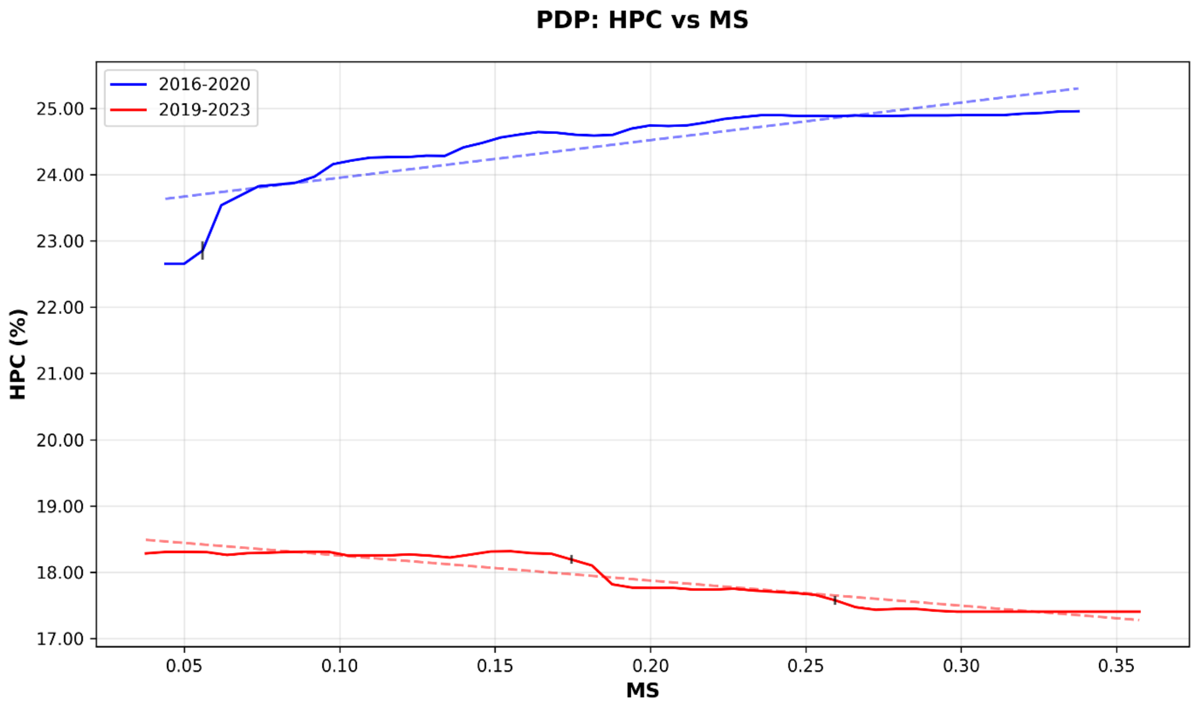

PDP: HPC vs FPR
HPC (%)
FPR
24.00
22.00
20.00
18.00
16.00
0.10
0.15
0.20
0.25
0.30
0.35
2016-2020
2019-2023

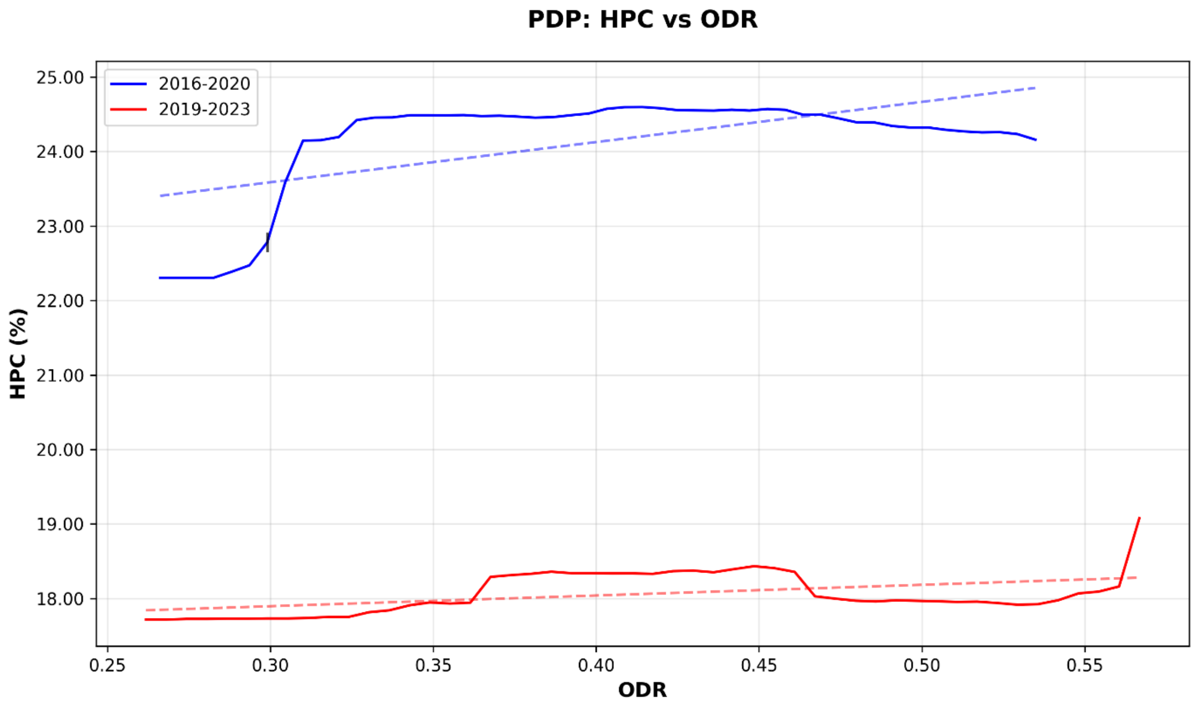

PDP: HPC vs GDP_pc
HPC (%)
GDP_pc
24.00
23.00
22.00
21.00
20.00
19.00
18.00
200
300
400
500
600
2016-2020
2019-2023

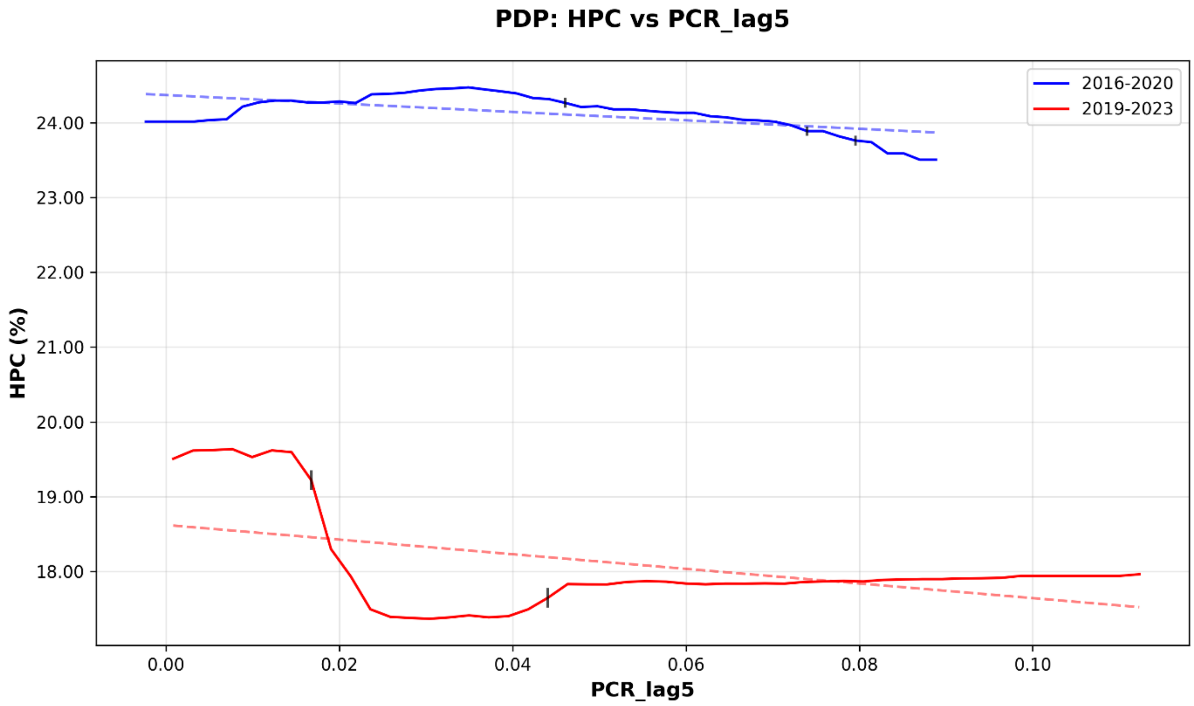

PDP: HPC vs HPC_lag3
HPC (%)
HPC_lag3
2016-2020
2019-2023
30.00
28.00
26.00
24.00
22.00
20.00
18.00
16.00
0.0
0.1
0.2
0.3
0.4

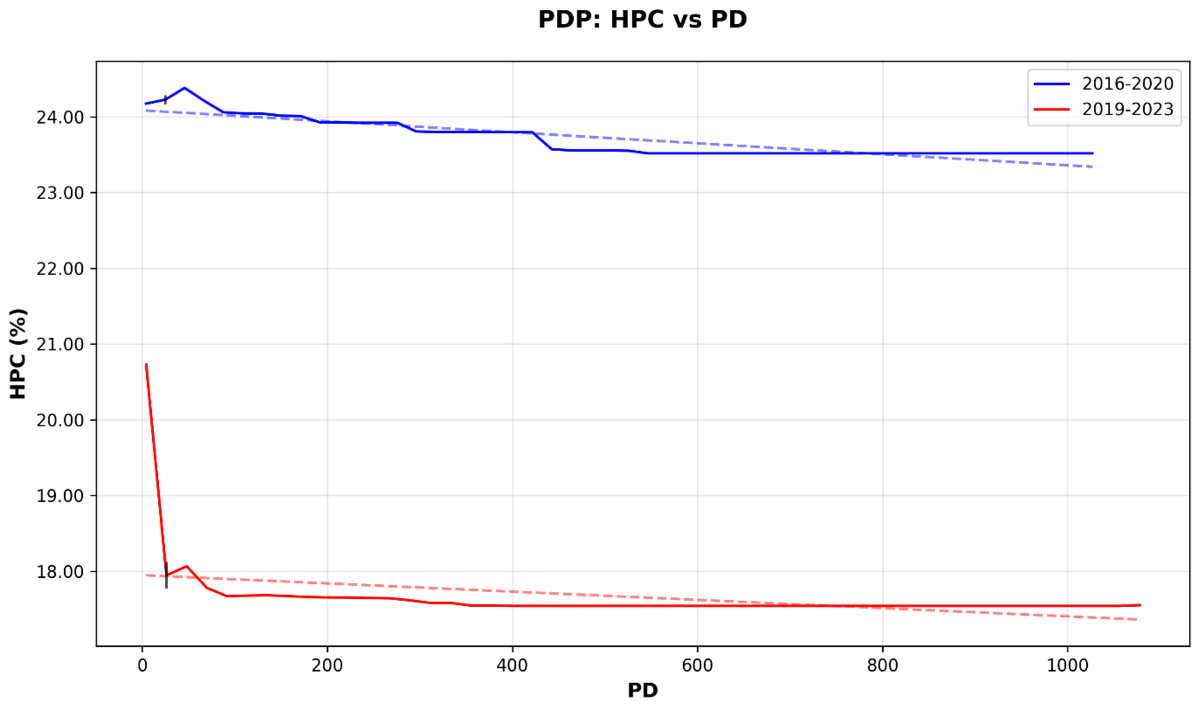

PDP: HPC vs INC
HPC (%)
INC
2016-2020
2019-2023
28.00
26.00
24.00
22.00
20.00
18.00
16.00
240
260
280
300
320
340
360
380

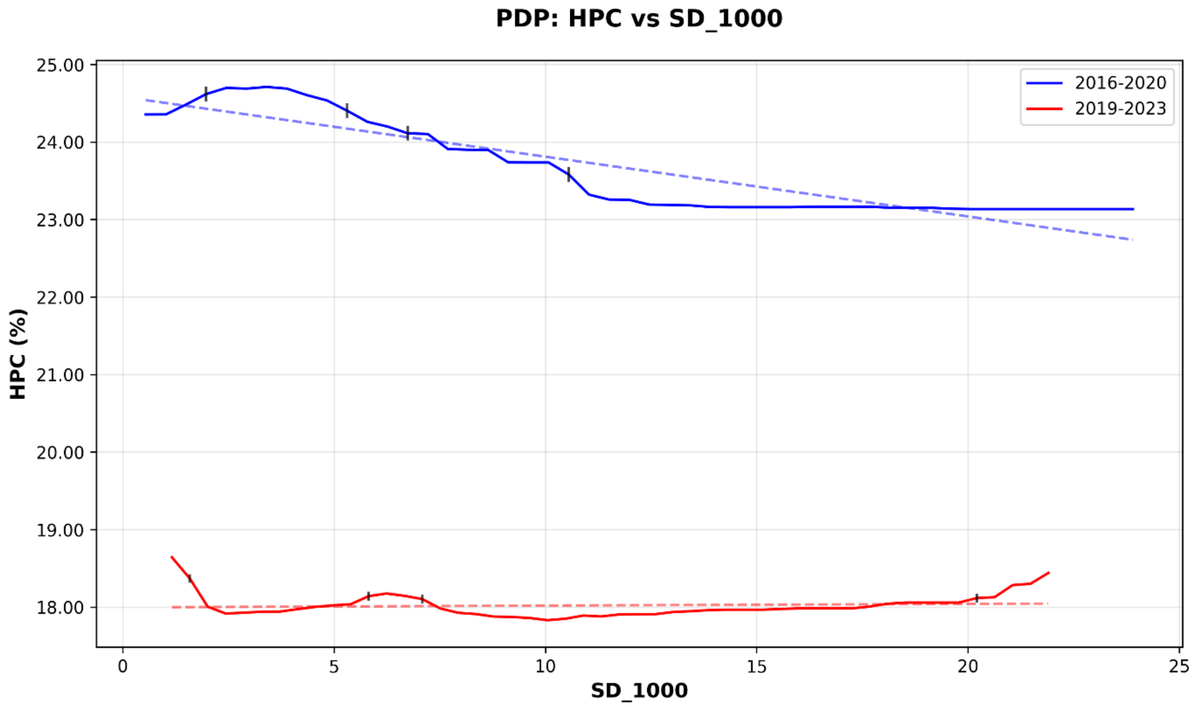

PDP: HPC vs INC_lag3
2016-2020
2019-2023
HPC (%)
INC_lag3
24.00
23.00
22.00
21.00
20.00
19.00
18.00
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0.18

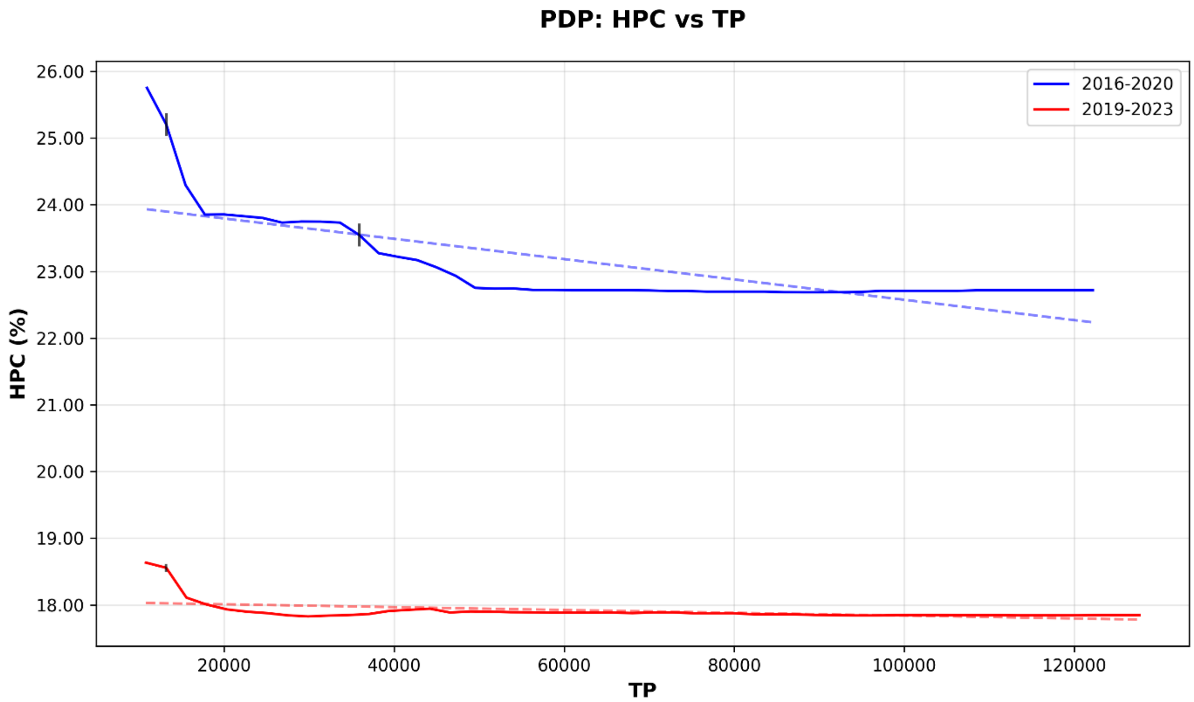

PDP: HPC vs MS
2016-2020
2019-2023
HPC (%)
MS
25.00
24.00
23.00
22.00
21.00
20.00
19.00
18.00
17.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35


PDP: HPC vs ODR
2016-2020
2019-2023
HPC (%)
ODR
25.00
24.00
23.00
22.00
21.00
20.00
19.00
18.00
0.25
0.30
0.35
0.40
0.45
0.50
0.55


PDP: HPC vs PCR_lag5
2016-2020
2019-2023
HPC (%)
PCR_lag5
24.00
23.00
22.00
21.00
20.00
19.00
18.00
0.00
0.02
0.04
0.06
0.08
0.10


PDP: HPC vs PD
2016-2020
2019-2023
HPC (%)
PD
24.00
23.00
22.00
21.00
20.00
19.00
18.00
0
200
400
600
800
1000


PDP: HPC vs SD_1000
2016-2020
2019-2023
HPC (%)
SD_1000
25.00
24.00
23.00
22.00
21.00
20.00
19.00
18.00
0
5
10
15
20
25


PDP: HPC vs TP
2016-2020
2019-2023
HPC (%)
26.00
25.00
24.00
23.00
22.00
21.00
20.00
19.00
18.00
20000
40000
60000
80000
100000
120000
TP

TRITA – ABE-MBT-25247
Stockholm, Sweden 2025

www.kth.se