Journal of LEADERSHIP AND MANAGEMENT

Volume 12 Issue 2
Published July 2026

Edited by

Sue Kong, Ph.D.

Adam Szpaderski, Ph.D.

Published by
Institute of Management and Leadership

JOURNAL OF LEADERSHIP AND MANAGEMENT

Volume 12, Issue 2
Published July 2026

Edited by

Sue Kong, Ph.D.

Adam Szpaderski, Ph.D.

Institute of Management and Leadership
A Program of the Health, Prosperity & Leadership Institute

HPL Publications

Institute of Management and Leadership

Institute of Management and Leadership is a program of
Health, Prosperity, and Leadership Institute

Publisher:

HPL Publications

PO Box 564, Douglassville, PA 19518, USA

Digital Version ISSN: 2391-6087

Print Version ISSN: 2998-1042

Print ISBN: 978-1-61305-087-3

Contents

Journal of Leadership and Management®
An International Journal of Strategic, Behavioral and Public Aspects of Leadership in Management
Cabell's Listed
A journal of the Institute of Leadership and Management Association, PO Box 564, Douglassville PA 19518. 484-332-3331. https://www.jleadershipmanagement.org

Quantifying the Value of Disciplined Execution in Relationship Marketing

Russell K. Lemken, East Carolina University
Cody Logan Chullen, East Carolina University, chullenc@ecu.edu
William J. Rowe, East Carolina University
John F. Kros, East Carolina University

Online published: July 2026 Print published: July 2026 Editor: Sue Kong	Authorship Roles and Conflict of Interest Statement is on file at the Journal of Leadership and Management offices, available on request. Contact editors@jleadershipmanagement.org

Abstract

Relationship marketing (RM) capabilities have long been recognized as important differentiators of firm performance. Yet assessments from industry and academia indicate not all RM strategies are equivalent and outcomes are often inconsistent. This study aims to explore the potential impact of disciplined execution during the implementation of new RM service strategies at the firm level – i.e., offering new relationship marketing approaches for current clients in the financial services industry. Adopting a resource-based perspective, combined with expert interviews, the authors consider *disciplined execution* to be a complementary resource consisting of five dimensions that collectively support how well a strategy is carried out (engagement, adoption, responsiveness, completeness, and tenacity). The authors use Bayesian inference to isolate the effects of different levels of disciplined execution across firms implementing RM service strategies. Research hypotheses are developed and tested using a unique dataset and quasi-experimental approach in collaboration with two leading financial services firms. The analysis shows that implementing RM service strategies positively influences annual recurring revenue, a focused measure of RM performance in financial services, and that higher levels of disciplined execution produce substantially greater performance gains.

KEYWORDS
Relationship marketing, Marketing capabilities, Resource-based theory, Disciplined execution, Bayesian analysis, Quasi-experiment

Acknowledgements: We note with sadness that our colleague Russell "Russ" Lemken passed away prior to the submission of this manuscript. He is remembered with respect and gratitude for his scholarly contributions. His role in this work is reflected in the authorship contributions statement.

Introduction

Relationship marketing (RM) has become a foundational construct for understanding how value is derived from relational exchanges. RM is used to improve customer acquisition, loyalty, and repeat purchase behavior, and RM capabilities have long been recognized as an important differentiator of firm performance (Evans & Laskin, 1994) and a source of competitive advantage (Day, 2000). Despite these benefits, not all RM service strategies achieve the expected outcomes (Zhang et al., 2016). Industry and academic assessments indicate that many RM service strategies fall short of performance expectations, with estimates of unmet expectations ranging from 69% to as high as 90% (Edinger, 2018; Taber, 2017). Thus, even with a robust RM literature stream, there remains a need to better understand the sources of variability in RM outcomes.

The current study addresses this need by examining the role of *disciplined execution* in RM service strategy effectiveness. Adopting a resource-based perspective and drawing on expert interviews in the financial services industry, we conceptualize disciplined execution as a complementary resource consisting of five dimensions related to strategy execution: engagement, adoption, responsiveness, completeness, and tenacity. This study advances prior research on competitive advantage and operating discipline by highlighting the mechanisms through which disciplined execution contributes to firm performance.

These mechanisms are often social and tacit in nature and embedded in the cultural fabric of the firm (March & Simon, 1958; Teece et al., 1997; Weick, 1979). They also depend on networks of strong and weak ties within the firm that are difficult for competitors to observe or imitate (Granovetter, 1973; 1985). As such, disciplined execution may represent a source of sustainable competitive advantage consistent with the imperfect imitability (I) and organizational (O) dimensions of marketing resources in the VRIO framework (Kozlenkova et al., 2014).

Using a unique dataset, we employ a two-step quasi-experimental approach in collaboration with two leading financial services firms. First, we compare the performance of firms that implemented new RM service strategies (i.e., offering new services to current clients) to firms that made no such strategic changes. Second, within the group of firms that implemented RM service strategies, we assess differences in disciplined execution. Change in annual recurring revenue (ARR) over the treatment period serves as our performance measure. We use a Bayesian multiple regression model to isolate the effects of disciplined execution across multi-disciplinary sales and service teams.

Our findings indicate that firms implementing new RM service strategies experienced increased ARR relative to the control group. Moreover, higher levels of disciplined execution were associated with significantly greater revenue gains. These results suggest that much of the benefit realized from implementing new RM service strategies depends on the level of disciplined execution during implementation. By demonstrating firm-level variability in disciplined execution and its impact on RM performance outcomes, this study identifies an important source of competitive marketing dynamics within a mature industry. The magnitude of these effects led participating firms to revise their RM service strategies to emphasize the five dimensions of disciplined execution identified in this study, further reinforcing the practical relevance of our findings.

The following sections develop our theoretical framework and hypotheses; describe our methodology, analysis, and results; and discuss the theoretical and practical implications of our study.

Resource-Based Theory and Disciplined Execution

The resource-based theory of the firm (RBT) has been an important foundation of marketing scholarship for more than 30 years (Barney, 1991; Barney et al., 2011; Slotegraaf et al., 2003; Vorhies & Morgan, 2005; Wernerfelt, 1984). Yet, despite the extensive use of RBT in marketing, Kozlenkova et al. (2014) argue that the process by which marketing resources are built, fine-tuned, and sustained over time remains inadequately explored. Empirical research, they contend, tends to focus primarily on the "valuable" (V) dimension of the four criteria for sustainable advantage, namely valuable, rare, imperfectly imitable, and organizationally exploitable, or VRIO. This tendency reinforces the view that RBT characterizes resources and capabilities as conferring advantage in a relatively fixed manner. Even so-called dynamic capabilities are often characterized as static in this sense (Barney & Clark, 2007).

Kozlenkova et al. (2014) further explore the evolution of RBT and evaluate RBT research in marketing using the VRIO framework as a basis for judging whether a marketing resource confers sustainable marketplace advantage (Barney & Hesterly, 2012). In describing gaps in RBT-based research in marketing, Kozlenkova et al. (2014, p.11) assert, “First, existing VRIO arguments are expressed mostly by citing past empirical research (Morgan et al., 2009), less frequently with the researcher’s own arguments (Evanschitzky et al., 2007), and almost never through an actual measurement of VRIO (Ramaswami et al., 2009). Second, the ‘O’ (organization) dimension of the VRIO framework is widely neglected.”

In the present study, we endeavor to address these theoretical gaps concerning marketing resources by using an empirical approach to formulate and test a set of hypotheses regarding the impact of disciplined execution on RM capabilities. In doing so, we examine how variation in disciplined execution across firms implementing the same RM service strategies shapes performance outcomes, thereby engaging the inimitable and organizational dimensions of the VRIO framework. By defining disciplined execution and evaluating its impact on a direct measure of RM results, we provide empirical insight into how organizational alignment and execution discipline contribute to sustainable advantage in marketing.

Understanding Disciplined Execution

We define "disciplined execution" as a strategic resource encompassing the dimensions of engagement, adoption, responsiveness, completeness, and tenacity. While execution and implementation are often used interchangeably in prior research, we distinguish them conceptually in this study. We use implementation to refer to the broader process of introducing and integrating new RM service strategies within the firm. In contrast, execution refers more specifically to the quality, consistency, and discipline with which those strategies are carried out behaviorally across teams and work processes. Disciplined execution therefore reflects how effectively firms enact RM service strategies during implementation.

Specifically, *engagement* involves the degree of active participation and effort put into implementing strategy. High engagement is demonstrated by interest, ownership, and involvement in the process of strategy implementation. *Adoption* refers to the extent to which the firm fully integrates and embraces the recommended practices and work processes, ensuring these processes are consistently applied across all relevant aspects of the operation. High adoption is demonstrated by full integration of the recommended practices into daily operations, consistent use of the prescribed work processes across team members, and visible alignment between stated strategy and actual behavior. *Responsiveness* refers to a willingness to accept coaching and input concerning adjustments to current work processes. High responsiveness indicates the ability to quickly and effectively act on feedback and suggestions for adjustments, adapting work processes to improve performance and align with evolving strategies. *Completeness* refers to the implementation and follow-through in using prescribed new approaches. High completeness is exhibited by thorough and sustained application of newly prescribed approaches, ensuring that all steps are carried out with attention to detail, and that any obstacles to full implementation are actively addressed. *Tenacity* involves a long-term commitment to implementing strategic changes and the perseverance to continue pursuing the successful implementation of these changes over time. High tenacity is demonstrated by commitment to the long-term goals of a strategy despite challenges or setbacks.

Based on expert interviews and our collaboration with industry partners during this study, these five components emerged as fundamental to successful implementation of RM service strategies and thereby form the underpinnings of disciplined execution. Our conceptualization of disciplined execution is consistent with the canon of RBT as reflected in a lengthy stream of economic, management, and marketing theory and research (Barney & Clark, 2007; Kozlenkova et al., 2014; March & Simon, 1958; Nelson & Winter, 1982; Weick, 1979; Wernerfelt, 1984). For discipline to have a sustained effect, collaborative work groups must have shared definitions-in-use and interpretative schemes for understanding customer input, common processes and technologies for accomplishing key outcomes, and established methods (Weick, 1979; Zablah et al., 2004). This shared framework is developed and modified over time to deliver specific business outcomes and can be disrupted by environmental changes and by changes in strategy (March & Simon, 1958). Consequently, observing the degree to which firms apply discipline in executing new RM service strategies provides a useful laboratory to test its effects in a defined period.

Effective implementation of new RM service strategies requires the members of the firm to invest a high degree of engagement in helping to define new strategies; hold a strong commitment to, and adoption of, the new practices identified; maintain a responsiveness to coaching; have a completeness in adhering to the totality of new procedures and methods; and possess tenacity in maintaining new practices throughout the stages of implementation. Taken together, these are the five dimensions that we tested and used to measure disciplined execution in the implementation of new RM service strategies. Figure 1 presents the research models guiding this study.

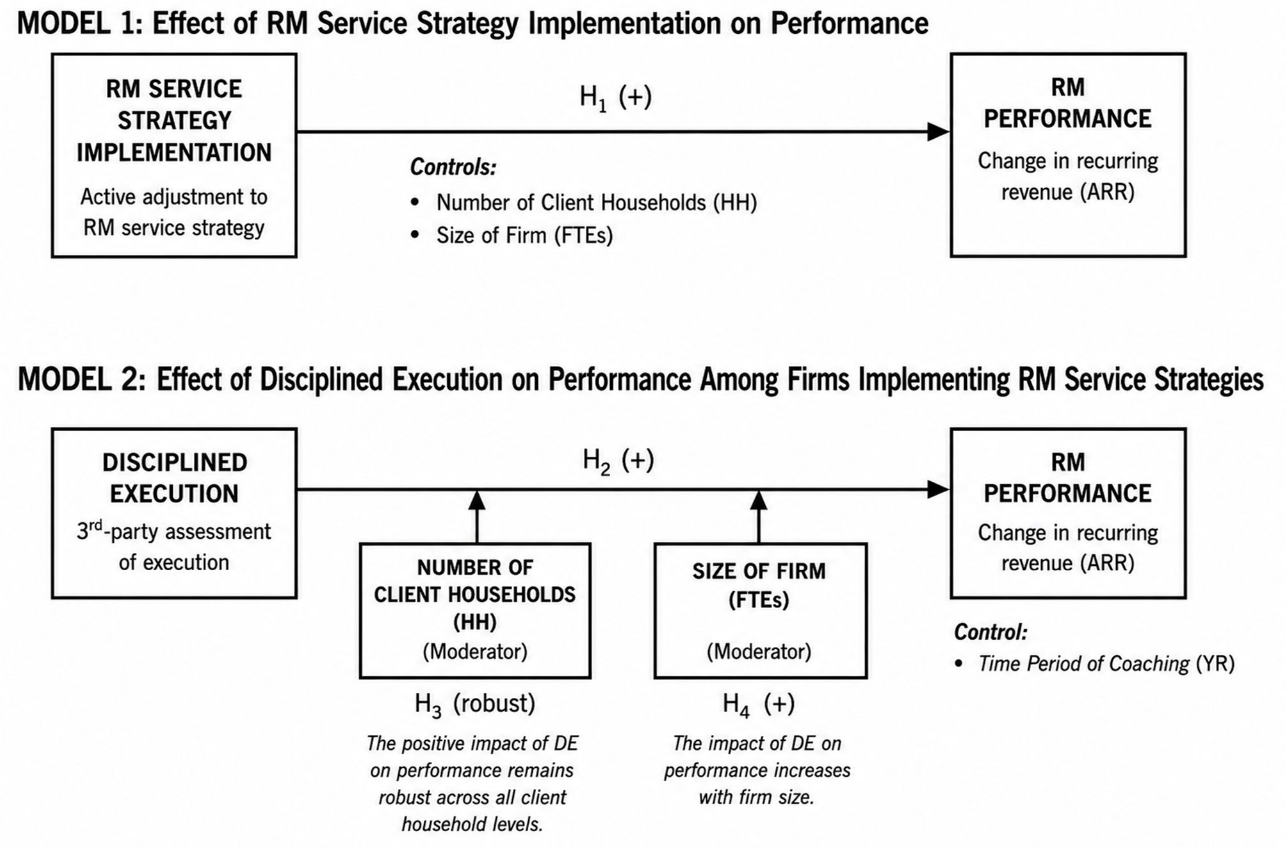


Figure 1. Research Models Evaluating RM Service Strategy Effectiveness and Disciplined Execution

Hypotheses

Relationship Marketing, Disciplined Execution, and Firm Performance

Establishing initial support for our study, we first hypothesize that firms implementing new RM service strategies will see a positive and significant increase in ARR, regardless of the level of disciplined execution (H_1). Next, using ARR as a performance outcome, we hypothesize that disciplined execution improves RM service strategy effectiveness by ensuring engagement, adoption, responsiveness to feedback, completeness in applying new methods, and tenacity in overcoming challenges. These dimensions collectively enable firms to realize the full potential of new RM service strategies, translating them into measurable performance gains. As such, we expect that firms with higher levels of disciplined execution will outperform those with lower levels.

Our conceptualization of disciplined execution is grounded in RBT, which suggests that firms gain competitive advantage through valuable, rare, inimitable, and organizationally exploitable resources (VRIO framework). While we posit that disciplined execution plays a role in successfully implementing RM service strategies, the five dimensions of disciplined execution—engagement, adoption, responsiveness, completeness, and tenacity—align with RBT's focus on organizational capabilities, making them difficult for competitors to replicate. This supports the view that disciplined execution is not only a complementary resource but may also be a driver of competitive advantage, as it enhances the firm's ability to implement and sustain strategic changes.

Based on the preceding logic, we expect that firms executing with greater disciplined execution should experience some performance benefit from these RM service strategies in the form of higher ARR. Further, we expect that these benefits will be magnified by increased disciplined execution. Perhaps more importantly, disciplined execution creates a defensible advantage, even compared to other firms that implement the same RM service strategies. Taken together, we hypothesize the following:

H_1: Firms that implement new RM service strategies will produce a positive and significant increase in ARR compared to a control group that does not implement new RM service strategies in the same period.

H_2: An increase in the disciplined execution of RM service strategies enhances the positive impact of RM service strategies on the firm's ARR.

Firm Size and Disciplined Execution

We use change in ARR to assess performance since it is a direct measure of the success of RM service strategy implementation. It includes the asset-based fees from long-term client relationships and excludes any additional sources of revenue for the firm, like commissions or other transaction-based income. In addition, financial advisory firms have a variety of client types that generate revenue, from those with very large assets to those with relatively modest wealth. Consequently, firms function with a wide range of clients. In the samples for our analytical models, the number of clients served by focal firms ranges from 77 to 908. By fostering engagement, adoption, responsiveness, completeness, and tenacity, disciplined execution ensures that RM service strategies are applied uniformly across diverse client segments, from high-net-worth individuals to smaller account holders. This universality of disciplined execution aligns with the principle that firm-specific capabilities can be leveraged to optimize outcomes across varied conditions and client profiles. Since all clients require advice and guidance, regardless of their asset level, disciplined execution can benefit firms implementing small account strategies as well as those serving very large asset accounts. Therefore, we predict:

H_3: The positive impact of disciplined execution on the RM-to-ARR relationship remains robust across all client sizes.

Finally, we anticipate that the size of the firm's employee base will have an impact on the influence of disciplined execution. The rationale for this belief is grounded in the idea that larger firms may face greater complexity in implementing RM service strategies due to the diversity of employees, functional areas, and operational processes. Larger employee bases often result in more significant coordination challenges leading to variability in execution. As such, disciplined execution becomes increasingly valuable in ensuring that RM service strategies are consistently implemented across the firm. The structured approach of disciplined execution provides a framework for aligning diverse teams and reducing variability, which is critical in larger, more complex organizations.

From an RBT perspective, the organizational dimension of the VRIO framework is particularly relevant for large firms. Larger firms typically have more resources, but these resources need to be mobilized and aligned effectively to achieve strategic objectives. Disciplined execution acts as an inimitable organizational capability that fosters alignment and consistency, amplifying the effectiveness of RM service strategies. Moreover, the scale of operations in larger firms allows them to fully capitalize on the benefits of RM service strategies in the presence of disciplined execution, resulting in a stronger RM-to-ARR relationship. In this context, disciplined execution serves as a critical lever to bridge the gap between strategic intent and performance outcomes, especially in firms with extensive employee bases. Therefore, we predict:

H_4: The impact of disciplined execution on the RM-to-ARR relationship increases with the size of the firm's employee base.

Methodology

Context of the Study

The context for this study is the U.S. financial services industry and the unit of analysis is the financial advisory practice (firm-level analysis). We define a financial advisory practice as a firm operating as a discrete business entity in a single location that serves a common set of clients. In this context, a firm may contain one or several licensed advisors and a staff of varying sizes and functions to support them. The financial services industry offers an excellent opportunity to study the impact of disciplined execution as firms in this industry are rich in collaborative sales and marketing teams implementing RM service strategies. Financial advisory firms have substantial autonomy in developing approaches to serve their clients. Consequently, they represent a wide range of sales and marketing configurations and operating philosophies. Also, since advisors must conform to strict regulatory prescriptions about how advice is delivered and the suitability of products, these firms are required to align sales and marketing approaches to provide consistent messages and direction based on sound defensible principles. This combination of industry conditions creates a wide variety of approaches within different firms with an industry-wide requirement for delivering basic service levels. This makes the financial services industry a robust laboratory for studying the performance effects of disciplined execution.

Financial advisory practices have become an increasingly recognized context for business scholarship. Per Law and Zuo (2021), “The U.S. Bureau of Labor Statistics (BLS) projected that the employment of financial advisors, which already accounted for 10% of total employment in the finance and insurance sector, would grow 15% from 2016 to 2026 – much faster than the average for all occupations (Bureau of Labor Statistics, 2018). Given the prevalence and importance of financial advisors, a growing literature has developed to further an understanding of their behavior (O’Hara, 2016).”

Data Collection

Developing our measure of disciplined execution, we first conducted directed interviews with 12 financial advisory industry experts to understand the common characteristics of firms and their respective sales and marketing teams with high levels of strategy execution (see Appendix A). Interview participants included trainers from large proprietary distribution systems, sales and marketing executives, and senior sales managers, among others. Specifically, the interviews were used to map discipline in execution across sales and service teams within firms and inform the development of the survey instrument provided in the appendix. The interviews focused on creating an accurate picture of the structures, work-processes, goal setting, and job descriptions used in financial advisory firms. The qualitative approach adopted for the interviews provided detailed descriptions of how teams in a firm work together to provide service to clients.

Qualitative data from the interviews was analyzed using NVivo 11 following an iterative process of categorization and comparison of the team activities, interactions, and outcomes uncovered in the interviews. The results were summarized and used to develop our initial measure of disciplined execution. The proposed measure and summaries of the interviews were then sent to and reviewed by the 12 industry experts originally interviewed. Their comments were incorporated in a second version of the disciplined execution measure. We used a 10-person subset of this expert group to pre-test the instrument. The results of the pre-test were used to refine our measure of disciplined execution prior to the primary survey.

Next, we used data collected from 30 financial advisory firms during the implementation of new RM service strategies over an eight-year period. The RM service strategies were developed and provided to the firms in our study by a large U.S. asset management firm that distributes individual investment products (e.g., mutual funds, managed accounts, and model portfolios) exclusively through licensed financial advisors. Supporting implementation of the new RM service strategies, the asset management firm provides specific direction and coaching to firms selling their financial products and services. Specific to the RM service strategies in our study, the coaching curriculum guides firms through the implementation of two new RM service strategies – one for managing small accounts more efficiently (referred to as small account servicing) and one for providing a higher level of value to larger clients with complex needs (referred to as key client management). Thirty financial advisory firms participated in the coaching curriculum supporting implementation of the new RM service strategies. Third-party assessments of disciplined execution were solicited from field consultants supporting the participating firms. We received 19 usable responses, which were used in the development and validation of the disciplined execution scale (see Table 1) and in the disciplined execution regression analysis (Model 2). In addition, these same 19 firms were compared against a control group of 23 firms that did not participate in the coaching curriculum to evaluate the broader impact of implementing RM service strategies (Model 1).

We use ARR as a performance measure. The firms in our study consistently report ARR in their accounting records so this data is both readily available and accurate. ARR is also commonly used as a metric of RM service strategy effectiveness in financial advisory firms. Note that revenue captured in ARR only includes regular fee income from long-term client relationships and thereby excludes transaction-based revenue, such as sales commissions or other one-time payments to the firm. These characteristics make ARR an ideal performance measure for our study.

The measure of change in ARR is an average for the years each firm was involved in implementing the new RM service strategies. The independent assessments of disciplined execution were provided by the asset management firm that supported the financial advisors responsible for implementing the new RM service strategies in the advisory firm. These ratings were solicited from the appropriate account service executives (i.e., coaches) after the focal financial advisors completed the coaching curriculum supporting RM service strategy implementation. As noted, 19 usable third-party execution-discipline ratings were obtained and used in the disciplined execution analysis (Model 2).

In order to compare firms that implemented new RM service strategies to a control group, we used a dataset with the financial results from 23 comparable financial advisory firms that did not implement new RM service strategies. The subset of 19 treatment firms with usable disciplined execution assessment data and complete analytical data was compared against 23 control firms that did not implement the RM service strategies, resulting in a total sample of 42 firms for the treatment-versus-control analysis (Model 1). The dataset for control firms was provided by a large broker/dealer organization that participated in the research process.

Structure of the Models

Two related analytical models were estimated in this study. Model 1 evaluated the overall impact of implementing RM service strategies by comparing the subset of treatment firms with complete analytical data (N = 19) against a control group of firms that did not implement the strategies (N = 23), resulting in a total sample of 42 firms. Model 2 evaluated the influence of differing levels of disciplined execution among the treatment firms with complete analytical data (N = 19). Third-party disciplined execution ratings used in scale development were obtained from 19 participating treatment firms (see Table 1), and the regression analyses relied on this same analytical sample.

In our model comparing firms that implement new RM service strategies to a control group (Model 1), we use a Bayesian approach to estimate parameters in a multiple regression that includes the number of full-time equivalent employees (FTE) and number of client households (HH) as covariates and a dummy variable for treatment to predict the change in ARR. In our model to evaluate the impact of disciplined execution (Model 2), we use a five-item, seven-point scale to solicit a third-party assessment of the [1] engagement level, [2] adoption level, [3] responsiveness to coaching and input, [4] completeness of implementation, and [5] tenacity of financial advisory teams that participated in the coaching curriculum (see Appendix A). We then use multiple regression to model the impact of different levels of disciplined execution on growth in ARR of the firms. Estimation is again carried out from a Bayesian perspective. Since ARR is generated by long-term, fee-based client relationships, it serves as an accurate measure of the effectiveness of RM service strategy implementation. The overall average rating of disciplined execution was 4.1 on the seven-point scale, with observed scores ranging from 2 to 6. The Cronbach's Alpha for the five-item measure of disciplined execution is 0.975, the component loadings for all items are between 0.946 and 0.968, and the construct accounts for 91% of the variation in the data (Cohen et al., 1983). Table 1 presents the correlation matrix for the disciplined execution scale items and demonstrates strong positive relationships across the five dimensions of the construct.

Table 1. Correlation Matrix for Disciplined Execution Scale Development Using Third-Party Ratings

		Engage	Adoption	Responsiveness	Completeness	Tenacity
Engage	Pearson Correlation	1	.890**	.913**	.874**	.878**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	19	19	19	19	19
Adoption	Pearson Correlation	.890**	1	.920**	.862**	.842**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	19	19	19	19	19
Responsiveness	Pearson Correlation	.913**	.920**	1	.900**	.884**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	19	19	19	19	19
Completeness	Pearson Correlation	.874**	.862**	.900**	1	.912**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	19	19	19	19	19
Tenacity	Pearson Correlation	.878**	.842**	.884**	.912**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	19	19	19	19	19

***Correlation is significant at the 0.01 level (2-tailed).*

Across the two models, we employed full-time equivalent employees (FTE), number of client households served (HH), and the time period of coaching (YR) as controls and covariates where appropriate. We used a dummy variable to control for the timing of the coaching and operating results in relationship to two years of high volatility in the capital markets. Our quasi-experimental framework provides an excellent opportunity to leverage the introduction and coaching in support of implementing new RM service strategies in advisory firms. The account service executive delivering the coaching then has the ability to monitor and gauge the level of adoption in each firm.

Bayesian methods provide an alternative to frequentist-based null-hypothesis testing that allows researchers to generate distributions for parameters and evaluate the sensitivity of analyses across varying levels of parameters (König & van der Schoot, 2018). The Bayesian paradigm also permits updates of parameter values and their associated distributions as new data become available (Gelman & Shalizi, 2013). When faced with relatively small datasets, like those used in this study, this feature of Bayesian methods is particularly useful. Even though we collected data over a period of eight years, we were able to update the analysis and provide insight and guidance to our business partners early and improve as more data became available. This addressed one important perennial disconnection in collaborations between academics and practitioners by providing value early in the process. We use vague priors for the model parameters in the initial models since there was no existing census-based evidence available on which to base informative priors (Albert, 2009). We follow a hypothetical-deductive approach to Bayesian inference, employing uninformative priors in our modeling that emphasize the data collected and the model's ability to test specific hypotheses outlined in the preceding section (Gelman & Shalizi, 2013).

We employed the Markov Chain Monte Carlo (MCMC) method to approximate the posterior distributions of all parameters by drawing very large representative random samples of the parameter values of interest in the two models. Using the very large sample of representative parameter draws, we computed the credible sets for parameter distributions and described their posterior distributions, including shapes. This in turn permits comparing values of different parameters in the two models (Kruschke et al., 2012). We use R statistical software and the JAGS (Just Another Gibbs Sampler) package to carry out all calculations related to the two Bayesian multiple linear regression models defined as follows: Model 1 assesses the treatment effect by regressing the change in ARR on FTE, HH, and a treatment dummy variable. Model 2 evaluates the impact of disciplined execution by regressing the change in ARR on the disciplined execution (DE) score, controlling for FTE, HH, the interaction between DE and FTE (DEFTE), and the coaching time period (YR).

For the control group comparison model (Model 1):

$p(y_i|x_{1i}, x_{2i}, x_{3i}, \beta_0, \beta_1, \beta_2, \beta_3, \sigma) = N(b_0+b_1x_{1i}+b_2x_{2i}+b_3x_{3i}, \sigma)$, where y_i is the change in ARR for the year, $x1_i$ is the number of full-time equivalent employees (FTE), $x2_i$ is the number of client households served (HH), $x3_i$ is the dummy variable indicating the treatment and control groups, and b_1, b_2, and b_3 are their respective coefficients. Table 2 reports the posterior correlation matrix for the control group model used in the Bayesian analysis.

For the disciplined execution model (Model 2):

$p(y_i|x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}, \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \sigma) = N(b_0+b_1x_{1i}+b_2x_{2i}+b_3x_{3i}+b_4x_{4i}+b_5x_{5i}, \sigma)$, where y_i is the change in ARR for the coaching curriculum period, x1i is the number of full-time equivalent employees (FTE), $x2_i$ is the number of client households served (HH), $x3_i$ is the measure of disciplined execution, $x4_i$ is the interaction of FTE and disciplined execution (DEFTE), $x5_i$ is a dummy variable representing the time period during which the firm received RM related coaching (YR), and b_1, b_2, b_3, b_4, and b_5 are their respective coefficients. Tables 3a and 3b report the posterior correlation matrices for the disciplined execution model used in the Bayesian analysis.

We used vague priors for the parameters associated with disciplined execution and DEFTE in the initial disciplined execution model to limit the influence of the choice of prior distributions on the posterior used for sampling. We used transformation to standardize the data prior to analysis, and so used normal priors. For the residual parameter of the model, we used a uniform prior with range of zero to 10, which is also very low in influence. In order to interpret the model, we retransformed the model parameters back to their original scale after modeling was complete. For each model, we initialized three chains at the mean values of the parameters and used 1,000 steps for tuning and 1,000 steps for burn-in and ran the full modeling process from beginning to end four different times and compared parameter values, plots, and diagnostics from each run. All four runs were essentially consistent with very small differences.

Table 2. Correlation Matrix of Posterior Values for the Control Group Model (Gibbs Sampler with 250,000 Steps)

	zsigma	zb0	zFTE	zHH	zTreat	sigma	b0	FTE	HH	Treat
zsigma	1.000									
zb0	0.005	1.000								
zFTE	-0.004	0.002	1.000							
zHH	0.002	-0.004	-0.607	1.000						
zTreat	-0.001	0.001	-0.011	-0.271	1.000					
sigma	1.000	0.005	-0.004	0.002	-0.001	1.000				
b0	0.005	0.476	-0.422	-0.253	-0.175	0.005	1.000			
FTE	-0.004	0.002	1.000	-0.607	-0.011	-0.004	-0.422	1.000		
HH	0.002	-0.004	-0.607	1.000	-0.271	0.002	-0.253	-0.607	1.000	
Treat	-0.001	0.001	-0.011	-0.271	1.000	-0.001	-0.175	-0.011	-0.271	1.000
Rsq	-0.004	0.000	0.736	0.022	-0.004	-0.004	-0.831	0.736	0.022	0.093

Table 3a. Correlation Matrix of Posterior Values for Disciplined Execution Model (Gibbs Sampler with 250,000 Steps)

	zsigma	zb0	zFTE	zHH	zDE	zDEFTE	zYR
zsigma	1.000						
zb0	0.003	1.000					
zFTE	0.007	0.003	1.000				
zHH	-0.003	-0.004	-0.160	1.000			
zDE	0.003	0.004	0.653	0.165	1.000		
zDEFTE	-0.006	-0.002	-0.934	-0.124	-0.765	1.000	
zYR	0.006	0.003	0.369	-0.230	0.138	-0.220	1.000
sigma	1.000	0.003	0.007	-0.003	0.003	-0.006	0.006
b0	-0.003	0.180	-0.695	-0.400	-0.872	0.794	-0.318
FTE	0.007	0.003	1.000	-0.160	0.653	-0.934	0.369
HH	-0.003	-0.004	-0.160	1.000	0.165	-0.124	-0.230
DE	0.003	0.004	0.653	0.165	1.000	-0.765	0.138
DEFTE	-0.006	-0.002	-0.934	-0.124	-0.765	1.000	-0.220
YR	0.006	0.003	0.369	-0.230	0.138	-0.220	1.000
Rsq	-0.004	0.001	-0.108	-0.075	-0.127	0.320	0.090

Table 3b. Correlation Matrix of Posterior Values for Disciplined Execution Model (Gibbs Sampler with 250,000 Steps) continued

	sigma	b0	FTE	HH	DE	DEFTE	YR
zsigma							
zb0							
zFTE							
zHH							
zDE							
zDEFTE							
zYR							
sigma	1.000						
b0	-0.003	1.000					
FTE	0.007	-0.695	1.000				
HH	-0.003	-0.400	-0.160	1.000			
DE	0.003	-0.872	0.653	0.165	1.000		
DEFTE	-0.006	0.794	-0.934	-0.124	-0.765	1.000	
YR	0.006	-0.318	0.369	-0.230	0.138	-0.220	1.000
Rsq	-0.004	-0.030	-0.108	-0.075	-0.127	0.320	0.090

We completed a series of diagnostic steps, including inspection of correlation matrices and scatterplots of posterior values, which produced no major causes for concern about autocorrelation, collinearity of variables or non-normal residual structures in the results of the models. We used the final iterations of each model for the purpose of reporting results and retained all the steps in the sampling process which totaled 250,000.

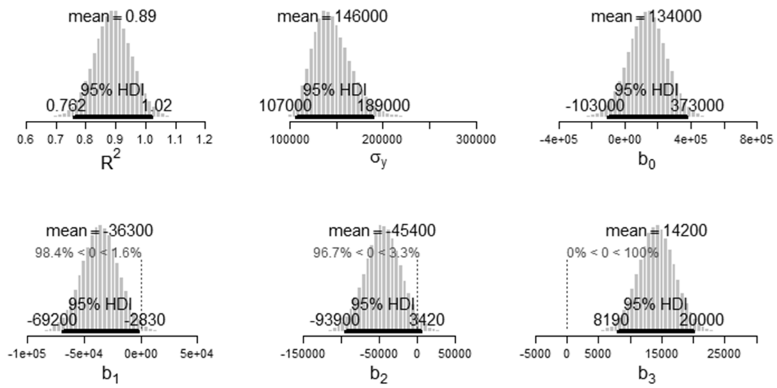
Since the disciplined execution data is provided by a third-party with active knowledge of the firms, the analysis and resulting models do not rely on self-reporting to assess disciplined execution. We consider this empirically grounded approach to measuring the impact of disciplined execution unbiased with regard to self-reporting and common method variance – two recurring concerns about behavioral methods used in RBT-based studies in marketing.

Results

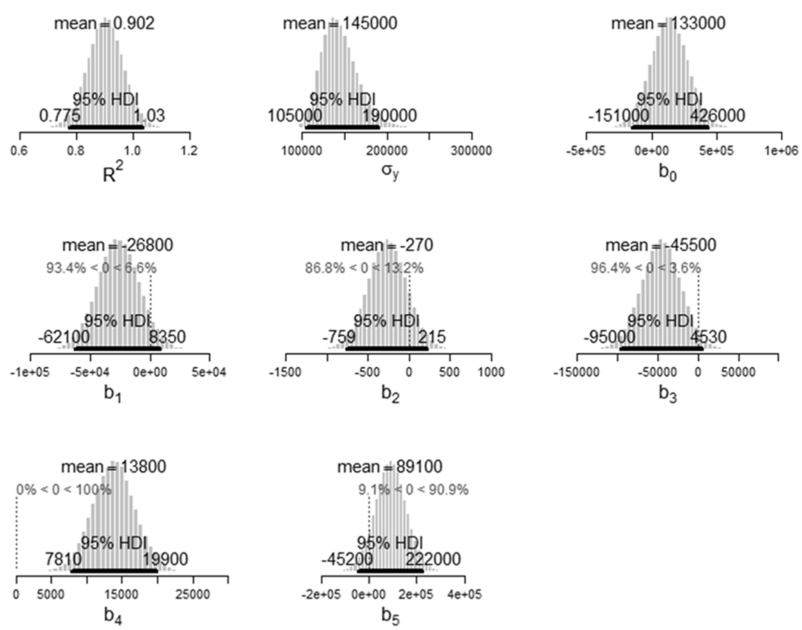
The results of the analysis provide strong empirical evidence of the positive effects of disciplined execution above and beyond electing to implement a new RM service strategy. By modeling and quantifying the difference in improvements to change in ARR among financial advisory firms with varying levels of disciplined execution, we provide a measure of direct impact on RM effectiveness. Figure 2 displays the results of the MCMC modeling using posterior distributions created by combining the non-influential prior distributions and the evidence from the sample data in the two models. The mean R2 for Model 1, comparing treatment and control firms, is 0.89, while the R2 for Model 2, evaluating disciplined execution within treatment firms, is 0.90.

Figure 2. Posterior Distributions for Model 1 (Control) and Model 2 (Disciplined Execution)

Model 1 – b_1 = FTE | b_2 = HH | b_3 = Treatment



Model 2 – b_1 = FTE | b_2 = HH | b_3 = DE | b_4 = DEFTE | b_5 = Time period



Tables 4 and 5 report the results of the control group and disciplined execution models used in the Bayesian analysis.

The analysis based on Model 1 to determine the effect of coaching compared to a control group shows that the average increase in ARR across the treatment-control comparison sample is $160,620. The treatment firms included in the analysis that received coaching have an increase in ARR of $190,877 compared to an increase of $131,177 for the control group, a difference of $59,700 or 46% attributable to the coaching. Model 2, measuring the effect of differing levels of disciplined execution among the 19 treatment firms with usable disciplined execution assessments and complete analytical data, had similar results. The average change in ARR among the 19 treatment firms included in the disciplined execution analysis is $191,340. The model shows that the average size firm with a mean rating for disciplined execution of 4.1 (on a scale of 7) has a change in ARR of $189,921. The same size firm with a disciplined execution score of 6 is modeled to have an increase in ARR of $267,173, a difference of $77,252 or 41%. Consequently, both H1 and H2 are supported.

Across the two analytical models, a firm of average size would create a benefit of $59,700 in change in ARR, on average, by adopting the RM service strategies taught in the coaching curriculum, according to Model 1. This alone was a very valuable proof point for the coaching team of the asset management firm. The same size firm that also increases its level of disciplined execution from average (4.1) to high (6) would produce an additional $77,252 in change in ARR, according to Model 2. Consequently, the results of the two models show that more than half (56.5%) of the overall gain available to an average-size financial advisory firm for the sample by both adopting new RM service strategies and increasing its ARR (from $131,177 to $267,173) is attributable to disciplined execution.

Table 4. Results of Control Group Model – Using Mean and 95% Credible Set

	DEPENDENT VARIABLE (CHANGE IN ANNUAL RECURRING REVENUE)			
	Control group	**Treatment group**	**Difference w/ treatment**	**Percent difference**
Upper bound of 95% credible set in posterior distributions	$475,406	$650,406	$175,000	37%
Mean values in posterior distributions	**$131,177**	**$190,877**	**$59,700**	**46%**
Lower bound of 95% credible set in posterior distributions	$-207,936	$-260,836	$-52,900	-25%

Table 5. Results of Disciplined Execution Model – Using Mean and 95% Credible Set

	DEPENDENT VARIABLE (CHANGE IN ANNUAL RECURRING REVENUE)			
	Base case (average firm)	**+2 DE in base case**	**Gain (loss) w/+2 DE**	**Percent gain (loss)**
Upper bound of 95% credible set values in posterior distributions	$1,445,007	$1,691,376	$246,369	17%
Mean values in posterior distributions	**$189,921**	**$267,173**	**$77,252**	**41%**
Lower bound of 95% credible set values in posterior distributions	$-1,057,471	$-1,147,534	$-90,063	-9%

The analysis demonstrates that increasing disciplined execution has a profound and positive influence on firm performance (ARR). Since firms rarely need to make capital investments to increase disciplined execution, these gains represent substantial increases in performance based on changes in behavior of the firm's advisory teams. This is particularly important for practitioners who struggle to explain and address the cause of weaker than expected results from RM service strategies. By offering a behavioral definition and measure of disciplined execution of RM service strategy implementation, we offer needed guidance for improving the success of all collaborative firm-level sales and marketing strategies.

We assessed multiple versions of both models using interaction terms between the independent variables and covariates to gauge their relationship and influence. We find that the distribution of parameters for client households (HH) and for the interaction of disciplined execution and households (DEHH) show they are likely to be near zero. For HH, the range of values for the parameter is $73.2\% < 0 < 26.8\%$ and for DEHH, the range is $48.2\% < 0 < 51.8\%$ in the model with all variables and interactions. Consequently, we find strong support for H3 that the number of client households is not a strong determinant of the effects of disciplined execution. The interaction term for the number of full-time equivalent employees (DEFTE) and disciplined execution is positive and significant in all versions of the model tested. In the final model for disciplined execution, the parameter for DEFTE has a mean value of 13,800, with a 95% credible set of 7,810 – 19,900, with a full distribution of $0\% < 0 < 100\%$. Therefore, we also find strong support for H4 since the effect of disciplined execution increases with FTEs.

Discussion

This study is the result of a collaborative effort between the authors and two large financial services firms searching for answers to an expensive and persistent problem in practice, the failure of relationship marketing strategies to achieve desired results. The direct experiences of the firms involved in this research, industry studies, and scholarly research concerning relationship marketing (RM) all agree that the implementation of new RM service strategies often falls significantly short of expectations. In our study, we propose and show that the ability to produce results through implementation of RM service strategies is tied closely to the level of discipline applied to the execution of new RM service strategies. We adopt a resource-based view of marketing capability development to define disciplined execution and provide a framework for measuring and applying disciplined execution to the implementation of RM service strategies.

Expert interviews conducted prior to experimentation helped to identify five dimensions of disciplined execution. The interviews also showed that firms with greater discipline were more flexible and able to adapt to difficult market conditions – a characteristic that should be considered in future studies of disciplined execution. Interview participants reasoned that disciplined firms have a reduced cognitive load due to having a more repeatable approach and benefit from a more stable baseline of firms to which they can make better comparisons when improvement opportunities are considered. And finally, they observed that firms with greater discipline are more successful at integrating new financial advisors into the firm as a growth strategy. These observations help to provide explanations of the mechanics at work that lead to performance gains in firms and therefore provide insight for practitioners.

As a result of this research, the two corporate partners working with the authors on this study altered their firm's coaching curricula to reflect the importance of team dynamics in implementing new RM service strategies. For example, one firm created a new practice management module concerning how to standardize certain workflows across advisory teams to increase consistency and efficiency. The other added multiple exercises for team building to enhance cohesion across the firm's sales and service teams. Both firms also indicated they have since changed their perspective on how to allocate resources within their own RM service strategy implementation efforts based on the importance of disciplined execution.

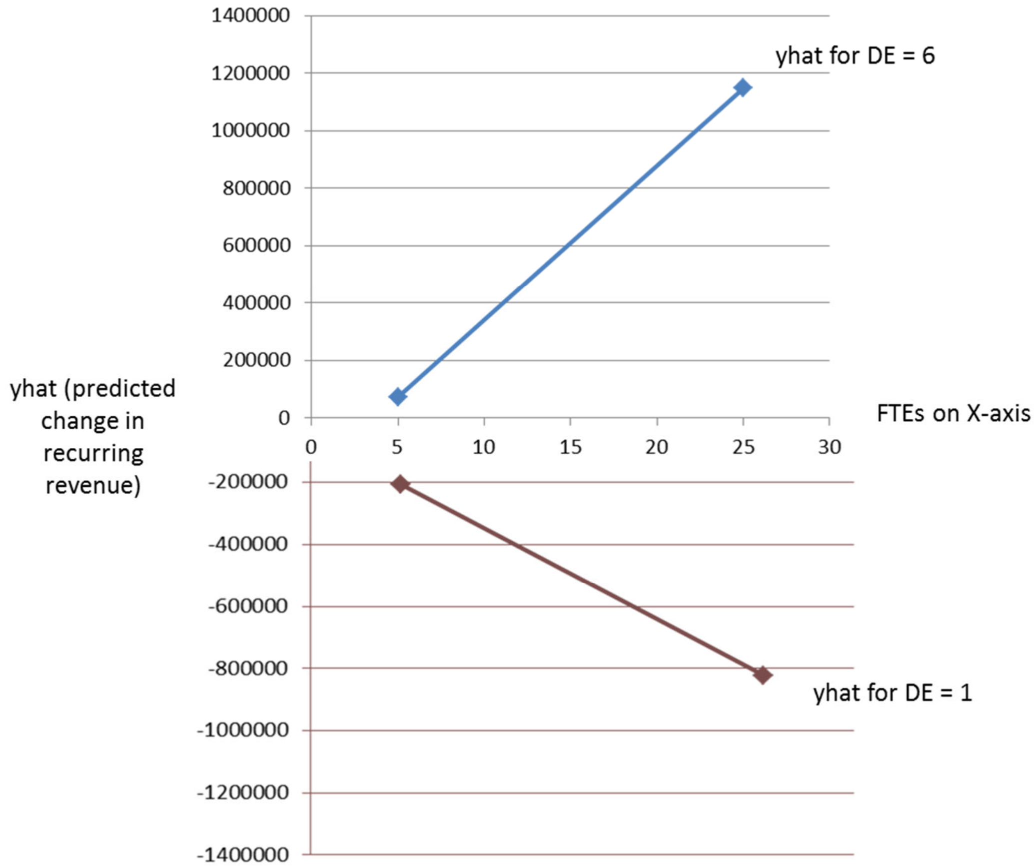


Figure 3. Changes in Recurring Revenue by Firm Size and DE Level

Our analysis provides strong empirical evidence that increasing discipline in execution has a significant and positive influence on firm performance. For a firm with average dimensions in the sample, increasing disciplined execution from average to high, as assessed by our measure, is worth $77,252 in additional ARR. This effect is robust across different sizes of client groups and is enhanced as the size of the employee base of the firm grows. Figure 3 illustrates the change in ARR across firm-size and discipline levels and highlights the effects of execution on performance.

Where high disciplined execution levels show gains across firm size, low levels have significantly negative impacts on both large and small firms. The two models used in our study provide support for all four research hypotheses concerning the influence of disciplined execution on the performance of RM service strategies. The models provide evidence that even firms receiving the exact same coaching in implementing RM service strategies can have divergent results based on their level of disciplined execution. Furthermore, the advantage attributable to disciplined execution should prove sustainable since it is based on inimitability and organizational factors related to the cognitive and behavioral dimensions of sales teams across the firm. These findings make an important and unique contribution to theory by showing a new path through which firms build sustainable competitive advantage by providing evidence of how inimitability and organizational factors influence the sustainability of relationship marketing resources.

Theoretical Implications

We contribute to scholarship in three ways. First, we define disciplined execution and develop a measure to assess the disciplined execution of RM service strategy implementation by identifying the behavioral foundations of multi-disciplinary sales and marketing execution. By specifying engagement, adoption, responsiveness, completeness, and tenacity as observable dimensions, we treat disciplined execution as a measurable organizational capability rather than an abstract managerial aspiration. Second, we use a unique dataset and quasi-experimental design to model the impact of disciplined execution on a direct measure of RM success, annual recurring revenue. By separating disciplined execution from the mere decision to implement RM service strategies, we distinguish between strategic intent and strategic realization. Our results show that performance differences often reflect variation in how consistently firms execute identical strategies. Execution is therefore not a secondary implementation issue, but a central driver of marketing performance. Third, we extend resource-based theory by examining the organizational dimension of the VRIO framework empirically. While prior research emphasizes valuable and rare resources, the organizational dimension has received far less direct attention. By operationalizing disciplined execution, we demonstrate that the "O" in VRIO can be understood as embedded behavioral routines and shared norms that govern how strategy is enacted across the firm. Organizational support is not simply structural alignment; it is disciplined, collective action sustained over time.

More broadly, our findings suggest that disciplined execution functions as a behavioral mechanism through which marketing capabilities translate into performance. Firms exposed to the same strategic prescriptions and coaching interventions produce different results depending on how systematically they execute. In mature industries where strategic options are widely known, disciplined execution may be a more important source of sustainable differentiation than strategy selection alone. Our results also inform work on dynamic capabilities. Discipline is often viewed as constraining, yet our findings indicate that consistent execution may enhance adaptability. By stabilizing routines and reducing unnecessary variability, disciplined firms free attention and coordination capacity for strategic adjustment. In this sense, discipline enables flexibility rather than restricting it.

Finally, we contribute methodologically by combining independent third-party behavioral assessments with Bayesian quasi-experimental modeling. This design addresses concerns about common method bias and strengthens empirical claims about organizational capabilities. By quantifying the magnitude of execution effects and demonstrating their interaction with firm size, we clarify how disciplined execution amplifies the returns to strategic investment. Taken together, these contributions move disciplined execution from managerial rhetoric to a clearly defined, empirically supported organizational capability that explains why identical marketing strategies yield different performance outcomes.

Managerial Implications

Our research contributes to practice in several important ways. First, we identify the specific behaviors that underlie disciplined execution by examining multi-disciplinary sales and marketing teams operating across the firm. The five dimensions identified in this study, engagement, adoption, responsiveness, completeness, and tenacity, provide managers with a practical framework for evaluating the quality of strategy implementation. Rather than relying on general impressions of commitment or effort, managers can assess these dimensions directly and use them to diagnose where execution is breaking down.

Second, our findings suggest that managers often overemphasize selecting new relationship marketing strategies while underinvesting in the behavioral conditions required to execute them well. The magnitude of the effects observed in this study indicates that disciplined execution accounts for a substantial portion of the gains associated with RM service strategy adoption. For the average firm in our sample, more than half of the increase in annual recurring revenue associated with implementing new RM service strategies was attributable to differences in execution discipline. Firms that neglect execution discipline may therefore leave meaningful revenue unrealized.

Third, the results show that improving disciplined execution does not require major capital investment. Unlike new technology systems or structural reorganizations, disciplined execution is rooted in collective behavior, coordination, and follow-through. Underperformance in RM service strategy implementation efforts may stem less from technical limitations and more from inconsistent accountability, uneven adherence to processes, and weak reinforcement of standards. Managers seeking stronger returns should focus on reinforcing consistent routines, clarifying expectations across sales and service teams, and establishing mechanisms that promote responsiveness and sustained commitment.

Fourth, disciplined execution becomes more valuable as organizational complexity increases. Larger firms face greater coordination demands and more variability in implementation. In these environments, disciplined execution serves as a stabilizing force that reduces inconsistency and aligns teams around shared standards. Growth without disciplined execution may amplify variability and erode the intended benefits of new RM service strategies.

Finally, because disciplined execution reflects embedded routines and shared norms, leadership practices related to hiring, onboarding, and team integration matter. Firms expanding through growth or acquisition should consider whether new personnel align with established execution standards and whether integration processes reinforce disciplined behaviors. Communication alone is not enough. Disciplined execution requires visible reinforcement of expectations, structured feedback, and sustained attention to follow-through.

As a result of this research, the five dimensions of disciplined execution have been incorporated into the coaching curricula of the participating firms. Both corporate partners indicated that future RM service strategy implementation efforts will allocate greater time and attention to strengthening execution discipline within multi-disciplinary teams. By framing execution as a measurable and financially consequential organizational capability, this study provides managers with clear direction for improving the success of collaborative sales and marketing strategies.

Limitations and Directions for Future Research

This research benefits from a deep and well-validated investigation of the mechanisms of a single type of sales organization. While this provides valuable detail about the capabilities that drive competitiveness, it limits generalizability. The Bayesian multiple regression used in our study relies on performance data and independent assessments of routine discipline to evaluate the effects of disciplined execution, with the analytical model based on observations from 19 firms. As such, future studies would benefit from a larger sample and additional quantitative analysis exploring additional factors that may have an impact on disciplined execution, like technological intensity and product lifecycle. While ARR provides a focused and widely used measure of RM performance in financial advisory firms, it does not directly capture profitability. Firms implementing RM service strategies may incur costs associated with coaching, employee time commitments, and implementation efforts that are not reflected in ARR alone. Consequently, future research may benefit from examining the relationship between disciplined execution, revenue growth, and profitability outcomes simultaneously.

As a guideline for scholars, generalizability of our findings can be tested by studying the use of adaptive and operational routines in other technically skilled sales and service contexts. For instance, a study of sales organizations in value-added reselling, common in the computer software and hardware industries, could provide a view of our hypotheses in a business-to-business context with greater technological dependence and shorter product lifecycles. In addition, future studies of disciplined execution in RM service strategies may choose to focus on the causes of friction in multi-disciplinary teams that can lead to a breakdown of discipline. Such research could also identify the antecedents of greater cohesion and effectiveness in sales and marketing teams. Understanding these cognitive and behavioral dimensions will help explain why certain teams are more engaged, committed, responsive, thorough, and resilient in executing RM service strategies.

Conclusion

In this study, we examined why identical relationship marketing strategies often produce divergent performance outcomes across firms. Using a resource-based perspective, independent third-party assessments of disciplined execution, and Bayesian quasi-experimental modeling, we demonstrate that execution discipline materially shapes the returns to implementation of RM service strategies. Firms that adopted new RM service strategies increased ARR, but those that executed with higher levels of engagement, adoption, responsiveness, completeness, and tenacity realized substantially greater gains. More than half of the total improvement available to the average firm in our sample was attributable to disciplined execution rather than strategy adoption alone. These findings clarify that performance variability in RM service strategy implementation is not simply a function of strategic choice, but of how consistently and collectively strategy is enacted across the firm. By defining and quantifying disciplined execution as an organizational capability, this study reinforces the importance of the "O" dimension in the VRIO framework and highlights disciplined execution as a measurable, financially consequential driver of competitive advantage in relationship marketing.

References

Albert, J. (2009). *Bayesian computation with R* (2nd ed.). Springer.

Barney, J. B., & Hesterly, W. S. (2012). *Strategic Management and Competitive Advantage: Concepts and Cases*. Pearson.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99-120. https://doi.org/10.1177/014920639101700108

Barney, J. B., & Clark, D. N. (2007). Resource-Based Theory: Creating and Sustaining Competitive Advantage. Oxford University Press.

Barney, J. B., Ketchen, D.J., & Wright, M. (2011). The future of resource-based theory revitalization or decline? *Journal of Management*, *37*(5), 1299-1315. https://doi.org/10.1177/0149206310391805

Bureau of Labor Statistics. (2018). *Occupational outlook handbook: Personal financial advisors*. U.S. Department of Labor. https://www.bls.gov/ooh/business-and-financial/personal-financial-advisors.htm

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (2nd ed.). Psychology Press. https://doi.org/10.4324/9781410606266

Day, G. S. (2000). Managing Market Relationships. *Journal of the Academy of Marketing Science*, *28*(1), 24-30. https://doi.org/10.1177/0092070300281003

Edinger, S. (2018, December). Why CRM projects fail and how to make them more successful. *Harvard Business Review*. https://hbr.org/2018/12/why-crm-projects-fail-and-how-to-make-them-more-successful

Evans, J. R. & Laskin, R. L. (1994). The relationship marketing process: A conceptualization and application. *Industrial Marketing Management*, *23*(5), 439–452. https://doi.org/10.1016/0019-8501(94)90007-8

Evanschitzky, H., Baumgarth, C., Hubbard, R., & Armstrong, J.S. (2007). Replication research's disturbing trend. *Journal of Business Research*, *60*(4), 411-415. https://doi.org/10.1016/j.jbusres.2006.12.003

Gelman, A. & Shalizi, C. R. (2013). Philosophy and the Practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology, 66*, 8-38. https://doi.org/10.1111/j.2044-8317.2011.02037.x

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, *91*(3), 481-510. https://www.jstor.org/stable/2780199

Granovetter, M.S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6) 1360-1380. https://www.jstor.org/stable/2776392

König C., & van der Schoot, R. (2018). Bayesian statistics in educational research: a look at the current state of affairs. *Educational Review, 70*(4), 486-509. https://doi.org/10.1080/00131911.2017.1350636

Kozlenkova, I.V., Samaha, S.A., & Palmatier, R.W. (2014). Resource-based theory in marketing. *Journal of the Academy of Marketing Science, 42*(1), 1-21. *https://doi.org/10.1007/s11747-013-0336-7*

Kruschke, J.K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods, 15*(4), 722-752. https://doi.org/10.1177/1094428112457829

Law, K.K.F., & Zuo, L. (2021). How does the Economy Shape the Financial Advisory Profession? *Management Science, 67*(4), 2466-2482. http://dx.doi.org/10.2139/ssrn.3277533

March, J. G., & Simon, H. A. (1958). *Organizations*. Blackwell Business.

Morgan, N.A., Vorhies, D.W., & Mason, C.H. (2009). Market orientation, marketing capabilities, and firm performance. *Strategic Management Journal*, *30*(8), 909-920. https://doi.org/10.1002/smj.764

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Belknap Press of Harvard University Press.

O'Hara, M. (2016). *Something for nothing: Arbitrage and ethics on Wall Street*. W. W. Norton & Company.

Ramaswami, S.N., Srivastava, R.K., & Bhargava, M. (2009). Market-based capabilities and financial performance of firms: Insights into marketing's contribution to firm value. *Journal of the Academy of Marketing Science, 37*(2), 97-116. http://doi.org/10.1007/s11747-008-0120-2

Slotegraaf, R.J., Moorman, C., & Inman, J.J. (2003). The role of firm resources in returns to market deployment. *Journal of Marketing Research, 40*(3), 295-309. https://www.jstor.org/stable/30038860

Taber, D. (2017, September). What to do when your CRM project fails. *CIO Magazine*, 61–62.

Teece, D.J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal, 18*(7), 509-533. https://doi.org/10.1142/9789812796929_0004

Vorhies, D.W. & Morgan, N.A. (2005). Benchmarking Marketing Capabilities for Sustainable Competitive Advantage. *Journal of Marketing, 69*(1), 80-94. https://doi.org/10.1509%2Fjmkg.69.1.80.55505

Weick, K. E. (1979). *The social psychology of organizing* (2nd ed.). McGraw-Hill.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal, 5*(2), 171-180. https://doi.org/10.1002/smj.4250050207

Zablah, A.R., Bellenger, D.N., & Johnston, W.J. (2004). Customer Relationship Management Implementation Gaps. *Journal of Personal Selling and Sales Management, 24*(4), 279-295. https://doi.org/10.1080/08853134.2004.10749038

Zhang, J. Z., Watson IV, G. F., Palmatier, R. W., & Dant, R. P. (2016). Dynamic Relationship Marketing. *Journal of Marketing*, *80*(5), 53-75. https://doi.org/10.1509/jm.15.0066

Appendix A. Assessments of Advisory Firm Disciplined Execution Level

(Cronbach's alpha = 0.975, Eigenvalue = 4.55, construct accounts for 91% of variation)

Using the questionnaire below, we surveyed the Account Service Executives (ASE) that delivered the coaching curriculum supporting the new RM service strategy to each firm in our study. The ASEs work for a third-party service provider and assist the firms' financial advisors with RM service strategy implementation. For clarity and shared understanding, the ASEs were provided descriptions for each facet of disciplined execution included in the assessment.

FOR ABC WEALTH ADVISORY INC.:

Based on your observations and experience working with this financial advisory firm, use the scale provided to compare this firm to other financial advisory firms you have assisted in the practice management program.

	Among the lowest			**Average**			**Among the highest**	
Measurement items:	1	2	3	4	5	6	7	**Component loading**
1. Engagement level in the practice management program	☐	☐	☐	☐	☐	☐	☐	.955
2. Full acceptance of recommended strategy and processes	☐	☐	☐	☐	☐	☐	☐	.946
3. Orchestration of team efforts in adjusting roles and work processes	☐	☐	☐	☐	☐	☐	☐	.968
4. Committing needed time and resources in adapting work processes	☐	☐	☐	☐	☐	☐	☐	.953
5. Tenacity and long-term commitment to implementing the strategy	☐	☐	☐	☐	☐	☐	☐	.947

Journal of Leadership and Management®
An International Journal of Strategic, Behavioral and Public Aspects of Leadership in Management
Cabell's Listed
A journal of the Institute of Leadership and Management Association, PO Box 564, Douglassville PA 19518. 484-332-3331. https://www.jleadershipmanagement.org

Quantifying Electoral Enthusiasm: County-Level Indices of Partisan Mobilization in Pennsylvania

Guido Giuntini, West Chester University, ggiuntini@wcupa.edu
Nicholas P. Nicoletti, Alvernia University
David G. Rhoads, Independent Researcher

Online published: July 2026 Print published: July 2026 Editor: Sue Kong	Authorship Roles and Conflict of Interest Statement is on file at the Journal of Leadership and Management offices, available on request. Contact editors@jleadershipmanagement.org

Abstract

This paper introduces two new indices for measuring electoral performance at the sub-state level. We develop the Measure of Voter Enthusiasm (MVE) and the Voting Success Index (VSI). The MVE measures how well a party's candidate performs relative to its own registrants' turnout, using voters registered outside the two major parties — Libertarian, Green, and No Affiliation registrants, collectively termed "Other" — as the competitive benchmark. The VSI extends this by incorporating registration totals, capturing the degree to which a party's vote total exceeds or falls short of its registration base relative to the full Other registration pool. We apply both indices to the 67 counties of Pennsylvania using data from the 2024 Presidential election, paired with socioeconomic data from the Census Bureau's American Community Survey. Correlation analysis and parsimonious OLS models reveal that Democratic index values (MVED, VSID) are strongly conditioned by county-level socioeconomic characteristics, particularly household income, poverty rates, educational attainment, and the proportion of foreign-born women. Republican index values, by contrast, show a weaker or no socioeconomic gradient. The VSIR is essentially uncorrelated with standard demographic predictors, suggesting that Republican voting success displayed comparatively little county-level socioeconomic variation. This asymmetry is consistent with a mobilization hypothesis grounded in the stability of partisan identification, where the indices are interpreted as county-level indicators of differential turnout and net Other-registrant capture rather than direct evidence of partisan crossover. Influence diagnostics confirm that the central findings are robust to the exclusion of individual influential counties, and the asymmetric socioeconomic structure of Democratic versus Republican mobilization is consistent with surge-and-decline theory, suggesting the VSIR null finding may reflect 2024 Republican surge conditions rather than a permanent structural feature. The indices are designed as diagnostic electoral-performance tools computable directly from publicly available registration and election data, with applications for both academic comparative research and practical campaign resource allocation. The findings have practical implications for campaign resource allocation and contribute a replicable measurement framework applicable to other closed-primary states.

KEYWORDS

Voter mobilization, partisan turnout, election administration, campaign resource allocation, sub-national electoral analysis.

Introduction

Pennsylvania was among the most consequential states in the 2024 Presidential election. PA's 67 counties span a wide range of socioeconomic and demographic contexts — from dense, diverse urban centers to sparsely populated rural communities — making it a valuable setting for examining sub-state variation in electoral performance. This paper uses county-level voter registration, turnout, and election results from the 2024 Presidential contest, combined with socioeconomic data from the Census Bureau's American Community Survey, to examine county-level electoral-performance patterns across Pennsylvania. To accomplish this task, we introduce two new indices to measure a party's appeal beyond its registered base: the *Measure of Voter Enthusiasm* (MVE) and the *Voting Success Index* (VSI).

A common technique in political science is the creation of a model predicting individual vote choice using socio-demographic variables, where the dependent variable is respondents' vote choice (estimated via logistic or probit regression) and the independent variables are socioeconomic, political party, ideological, and policy-level characteristics. Scholars often use the American National Election Studies (2025) data for these individual-level models. An approach like this identifies which voter subgroups support specific candidates, and the results can be aggregated to describe group-level patterns. The socioeconomic variables used in the present analysis were selected based on initial post-election analysis of the 2024 results, which indicated that rural residents, young males, and those without a college degree were among the demographic groups that most heavily favored Trump (Sanders, 2024).

Existing research consistently finds that voter crossover is rare (Lewis-Beck & Martini, 2020; Skocpol & Fiorina, 1999). A very high percentage of party registrants vote for their registered party (Green, Palmquist, & Schickler, 2002). Partisanship, measured on the standard seven-point scale from strong Democrat to strong Republican (Elliott, 2023), functions as a stable social identity rather than a running evaluation of candidates or policies, and has historically been one of the best predictors of vote choice. While the proportion of voters who identify as "independent" appears large, most of these independents lean toward one party and vote nearly as reliably as weak partisans (Keith, Magleby, Nelson, Orr, Westlye, & Wolfinger, 1992). Only pure independents — those with no partisan lean — genuinely shift between parties in each election, and they constitute roughly 7 to 10 percent of the electorate and are also the least likely to turn out (Green et al., 2002). However, Dalton (2008) distinguishes between apolitical independents, who are disengaged from electoral politics, and apartisan citizens, who are politically informed and issue-oriented but deliberately unattached to either party. The latter group constitutes the genuinely competitive pool the MVE and VSI are designed to capture — voters whose ballots are available to either party and who are engaged enough to actually turn out. The implication is that winning elections is fundamentally about which party mobilizes its base more effectively and captures this competitive pool of genuinely unattached voters, rather than persuading partisans to cross over (Keith et al., 1992; Petrocik, 1981; Rosenstone & Hansen, 1993).

This mobilization hypothesis is central to the interpretation of the indices proposed in this paper. When MVE or VSI values vary across counties, the variation most likely reflects differences in partisan turnout rates and each party's success at attracting voters registered as "Other" — not systematic defection by registered Democrats or Republicans. This framing is consistent with the broader literature on partisan stability while allowing for the influence of economic perceptions at the margin. Lewis-Beck & Martini (2020) find that objective economic perceptions have a substantive impact on support for incumbent candidates, though this effect is attenuated when controlling for partisanship. However, a growing body of research suggests that economic perceptions are themselves shaped by partisan identity. Bartels (2002) demonstrates that partisans evaluate identical economic conditions differently depending on which party holds the presidency, and Evans & Andersen (2006) present evidence that the dominant causal direction runs from partisanship to economic perception rather than the reverse. If voters perceive the economy through a partisan lens, then even the economic voting channel reinforces existing loyalties rather than driving crossover. This dynamic is likely amplified in an era of increased polarization, media atomization, and the evolving information landscape (Kavanagh & Rich, 2018). Similarly, Green & McElwee (2019) find that racial attitudes explain substantial variation in white voter support for Trump, with economic distress adding explanatory power after controlling for those attitudes. However, racial attitudes themselves are strongly sorted along partisan lines (Tesler, 2016), reinforcing the view that these factors operate within rather than across partisan boundaries. This loyalty is further entrenched by affective polarization, in which partisan identity functions as a viscerally charged identity that amplifies in-group cohesion and out-group hostility, independent of policy disagreement, effectively insulating partisan voters from cross-cutting appeals (Mason, 2018; Kalmoe & Mason, 2022).

Unlike the standard individual-level models, this paper uses aggregate county-level data to examine group-level patterns. Therefore, it is important to acknowledge the risk of the ecological fallacy, which highlights the difficulty of generalizing individual voters from aggregated data (Subramanian, Jones, Kaddour, & Krieger, 2009). We do not claim to identify individual-level vote switching or the likelihood of individuals' voting propensities. Rather, the indices describe county-level mobilization patterns, and the socioeconomic correlates identify counties where each party's index values indicate stronger or weaker electoral performance relative to its registration base. General trends at the aggregate level remain useful for understanding and analyzing election outcomes, particularly for the practical purpose of informing campaign resource allocation.

The indices are designed to serve two complementary purposes. Academically, they provide a replicable measurement framework for examining sub-state variation in partisan mobilization that can be applied across election cycles and closed-primary states, enabling comparative research into when and where mobilization asymmetries emerge. Practically, they function as post-election diagnostic tools: a positive MVE or VSI in a given county indicates that a party's candidate outperformed what its registration base alone would predict, while a negative value flags a mobilization deficit — either insufficient base turnout, failure to capture Other-registrant votes, or both. Campaign organizations can use these diagnostics to identify counties where marginal investments in turnout operations are most likely to matter and to assess whether underperformance reflects structural disadvantage or addressable mobilization gaps. Both uses depend on the same core property: that the indices decompose electoral performance into constituent parts rather than collapsing it into a single vote share.

The paper proceeds as follows. The next section reviews existing voting indices in the political science literature. We then present the MVE and VSI indices and describe their construction. The data and descriptive statistics for all 67 Pennsylvania counties are presented below. We present a correlation analysis examining the bivariate relationships between the indices and socioeconomic characteristics, followed by parsimonious OLS regression models that confirm these patterns in a multivariate context. We conclude with a discussion of findings, limitations, and directions for future research.

Reviewing Current Indexes

Having established the theoretical basis for interpreting electoral performance as a function of mobilization rather than crossover, we turn to existing measures of voting and party success in the political science literature. Several indices address related dimensions of electoral performance, though none capture precisely what the MVE and VSI are designed to measure.

In the political science literature, voting success is often used as a measure of party cohesion in legislatures or as a legislative effectiveness score. These scores often gauge ideological cohesion in Congress and reflect polarization. For example, in a non-majoritarian system, Gallagher (1991) measures proportional representation in terms of seat share. In the United States, with single-member plurality, which results in a two-party system based on Duverger's Law (Golosov, 2024), Gallagher's measure doesn't often work.

Measures of legislative ideology, such as DW-NOMINATE scores (Poole & Rosenthal, 1985; McCarty, Poole, & Rosenthal, 2006), offer a rigorous method for tracking ideological polarization among elected officials over time, but they capture elite behavior after electoral outcomes rather than the voter mobilization dynamics that produce those outcomes in the first place.

Measures like incumbent party approval are sometimes used. Some of the presidential election prediction models use very basic regressions that include trial heat polling for the incumbent, economic growth, and the convention bump to predict the popular vote. Campbell's (2008) model, for instance, uses trial heat polling to predict the incumbent's popular vote percentage with good accuracy over time. In the past, measures of party effectiveness, such as indices of "party vigor" (Key, 1949), based on local competitiveness, organization, and voter participation, were employed, but they are less used today.

The Cook Partisan Voting Index

The Cook Partisan Voting Index (Cook PVI), first introduced in 1997, measures how each state and congressional district performs at the presidential level compared to the nation as a whole (Cook Political Report). The index is displayed as a letter-plus-number combination (e.g., D+5 or R+12), with the letter indicating the party that outperformed and the number showing the margin above the national average. The Cook PVI can be applied at the local level and has been found to be a strong predictor of vote choice regardless of controls. Paul et al. (2022) use the index to examine 16 health outcomes for U.S. children, finding that Democratic-leaning states had better median outcomes across all 16 measures.

While the Cook PVI is a useful heuristic for describing district-level partisan lean, it differs from the MVE and VSI in three important respects. First, the Cook PVI benchmarks performance against a national average across two election cycles, making it a measure of relative partisan geography rather than within-election mobilization. Second, it does not decompose performance into its constituent parts — turnout of the registered base versus capture of Other-registered voters — which is precisely the distinction the MVE and VSI are designed to make. Third, because the Cook PVI is anchored to presidential vote share rather than voter registration, it cannot distinguish between a strong performance driven by exceptional base turnout and one driven by Other-voter capture. The MVE and VSI, by contrast, operate within a single election and measure electoral outcomes directly against the registration base, allowing researchers and practitioners to identify why a party overperformed or underperformed in each county rather than simply that it did.

The Partisan Turnout Rate

Gimpel & Schuknecht (2003) propose a Partisan Turnout Rate aimed at measuring impediments to voting:

$$\text{Partisan Turnout Rate} = \frac{\text{Votes cast for party i}}{\text{Registered voters for party i}}$$

Their dependent variable is the percentage of registered voters who cast ballots in the 2000 presidential election in each precinct. This index shares some features with the indices we propose, though it does not isolate the Other-voter capture component.

The Turnout Overperformance Index

Biggers & Hanmer (2017) use a Turnout Overperformance Index:

$$\text{Turnout Overperformance} = \frac{\text{Votes for party i} - \text{Registered party i voters}}{\text{Total registered voters}}$$

They employ this measure to examine the effects of voter identification laws, finding that the propensity to adopt such laws is greatest when control of the governor's office and legislature switches to Republicans, and that this likelihood increases as the size of Black and Latino populations in the state expands. While structurally similar to the VSI in using total party registration rather than turnout in the numerator, the Turnout Overperformance Index normalizes against total registered voters rather than the Other-registrant pool, conflating what is the theoretical competitive electorate with the full registration base and obscuring the theoretically distinct contribution of unaffiliated and minor-party voters to each party's electoral outcome.

Temporal Indices of Turnout and Vote Share

Several indices measure electoral effectiveness across time rather than within a single election. Fraga (2018) uses a Turnout Change Index comparing current and previous election turnout. Sides, Tesler, & Vavreck (2018) use a Change in Partisan Vote Share comparing a party's vote share across consecutive elections. Berinsky (2005) uses a Voting Rate index measuring the effectiveness of early voting mobilization by party. These temporal indices capture important dynamics but require multi-election data and do not decompose within-election mobilization patterns.

Specifically, Fraga (2018) uses a Turnout Change Index,

$$\text{Turnout change} = \frac{\text{Turnout in current election}}{\text{Turnout in previous election}}$$

Sides et al. (2018) use an index called Change in Partisan Vote Share,

$$\text{Change in partisan vote share} = \frac{\text{Vote share for party i in current election}}{\text{Vote share for party i in previous election}}$$

Berinsky (2005) uses a Voting Rate index to measure a party's effectiveness in mobilizing early voting.

$$\text{Voting rate} = \frac{\text{Early votes cast for party i}}{\text{Registered voters for party i}}$$

Some studies examine election result swings in a polarized landscape (Shaw, 2012), and others use district-level socio-demographic data and partisanship measures (Levendusky, Pope, & Jackman, 2008). Levendusky et al. note a significant concern with this approach:

> *"A measure of district partisanship that relied solely on demographic characteristics of the district suffers from an obvious threat to validity. Demographic attributes are generally considered antecedents of partisanship, rather than indicators of it. So, while demographic characteristics may correlate highly with one another and would appear to measure something about districts, there is no guarantee that demographic characteristics alone would permit us to locate congressional districts on a partisan continuum." (pp. 736)*

This concern is well-taken. The present paper does not attempt to infer county-level partisanship from demographics. Rather, we propose indices that directly measure observable electoral outcomes — the gap between a party's vote total and its registration base — and then ask which socioeconomic characteristics are associated with variation in those outcomes. The question is practical: in a world of limited resources where the precise vote of each individual is unknown, can we identify county-level patterns in mobilization and independent capture that would inform campaign strategy? The MVE and VSI indices are designed to answer that question. Unlike the Cook PVI, they decompose performance relative to registration rather than comparing it to a national baseline. Unlike the temporal indices, they operate within a single election. And unlike purely demographic approaches, they measure an electoral outcome directly rather than inferring partisanship from covariates.

The distinctions outlined above are not merely technical. Each existing index embeds a theoretical choice about what constitutes the relevant competitive electorate — the national average, the prior election baseline, or the full registration pool — and that choice shapes what the index can and cannot reveal about mobilization. Table 1 summarizes these distinctions and positions the MVE and VSI relative to the existing landscape.

Table 1. Comparison of Existing and Proposed Electoral Performance Indices

Index	What It Measures	Key Assumption	Unique Analytical Leverage
Cook Partisan Voting Index (PVI)	Partisan lean relative to the national average, averaged across two presidential elections	National average vote share is the appropriate benchmark; partisan geography is stable across cycles	Identifies structurally partisan districts; strong predictor of vote choice; does not decompose within-election mobilization
Partisan Turnout Rate (Gimpel & Schuknecht, 2003)	Share of registered party-*i* voters who cast ballots	Registered partisans are the relevant electorate; turnout is the primary mobilization signal	Captures base mobilization rate; does not isolate Other-voter capture
Turnout Overperformance Index (Biggers & Hanmer, 2017)	Net votes for party-*i* beyond its registration base, normalized against total registered voters	Total registration is the appropriate denominator	Captures net over- or underperformance; conflates Other-voter contribution with base turnout effects
Temporal Indices (Fraga, 2018; Sides et al., 2018)	Change in turnout or vote share across consecutive elections	The prior election is the appropriate baseline	Captures electoral trends over time; requires multi-election data; cannot decompose within-election patterns
Measure of Voter Enthusiasm (MVE) (proposed here)	Net votes beyond party-*i* registrant turnout, normalized against Other-registrant turnout	The competitive electorate consists of unaffiliated and minor-party voters; registered partisan defection is negligible	Isolates Other-voter capture within a single election; separable from base turnout effects
Voting Success Index (VSI) (proposed here)	Net vote performance relative to party-*i* registration base, normalized against Other-registrant registration	The competitive electorate consists of unaffiliated and minor-party voters; partisan crossover is negligible	Captures registered base mobilization and Other-voter capture simultaneously; enables within-election asymmetric partisan comparison

The Indices

Before formally presenting the indices, we clarify their conceptual status. The Measure of Voter Enthusiasm and the Voting Success Index are diagnostic electoral-performance measures — arithmetic functions of publicly available registration, turnout, and vote totals — rather than direct measures of psychological motivation or voter affect. We use “enthusiasm” as a shorthand for a party's mobilization performance relative to its registration base. That is the degree to which a party's candidate attracted votes beyond what its own registrants' turnout would predict, benchmarked against the unaffiliated and minor-party registrant pool. This framing is intentional. Because the indices are constructed from aggregate county-level data, they cannot establish individual-level psychological states; they can only determine whether a party's electoral outcome exceeded, matched, or fell short of what its registration base and turnout would mechanically imply. Index variation across counties, therefore, is interpreted as evidence of differences in electoral performance relative to base-turnout and Other-registrant benchmarks — the two observable electoral processes the mobilization hypothesis identifies as the primary determinants of election outcomes — rather than unmeasured differences in voter motivation. The denominator is restricted to Other-registrant ballots rather than all non-party-*i* ballots, reflecting the theoretical assumption, supported by the partisan stability literature, that the genuinely competitive electorate consists of unaffiliated and minor-party voters rather than registered partisans who might cross over. This assumption is empirically grounded in the 2024 Pennsylvania context, as third-party and independent candidates received approximately 0.97% of total votes statewide, confirming that Other-registrant ballots flowed overwhelmingly to the two major-party candidates (Pennsylvania Department of State, 2024).

To measure county-level electoral-performance patterns associated with socioeconomic context, we propose two indices with slightly different aims. One is the Measure of Voter Enthusiasm (MVE), the other is the Voting Success Index (VSI). The Measure of Voter Enthusiasm (MVE) is calculated as follows:

$$\mathrm{MVE} = \frac{\text{Votes cast for party i} - \text{registered party i ballots cast}}{\text{Registered Others ballots cast}}$$

The index is positive if a party's candidate receives more votes than the number of registered voters in that party who cast ballots, consistent, under the low-crossover assumption, with net vote support beyond the party's own registrant turnout. A negative MVE value indicates net underperformance relative to the party's own registrant turnout. A value greater than 1 indicates that the party's net vote surplus beyond its own registrant turnout exceeds the number of Other ballots cast — an upper bound that flags exceptionally strong performance among unaffiliated and minor-party voters. The denominator is restricted to Other ballots cast rather than all non-party-*i* ballots cast, reflecting the theoretical assumption, grounded in the partisan stability literature, that registered partisans rarely defect to the opposing party (Green, Palmquist, & Schickler, 2002; Keith et al., 1992). Under conditions of high partisan polarization — such as the 2024 Presidential contest — this assumption is particularly well-supported, as partisan defection rates have been documented at historically low levels in recent election cycles (Bartels, 2000). The genuinely competitive pool consists of unaffiliated and minor-party registrants. An additional assumption, consistent with the 2024 Presidential election results, is that independent and third-party candidates received a negligible percentage of total votes.

As the data will show, the MVE calculated for Republicans across different geographic units tends to have a strong negative correlation with the MVE calculated for Democrats, since net positive swings in independent votes in one party's direction must be offset by negative swings in the other direction.

The main limitation of the MVE is that it does not incorporate turnout, as it measures only ballots cast. Differential turnout rates between parties can cause the index to omit important information. Consider two counties with identical registration splits and identical vote shares. If one county had 90% Republican turnout and 70% Democratic turnout, while the other had equal turnout from both parties, MVE would not capture this difference.

For this reason, we also propose the Voting Success Index (VSI), which incorporates registration, turnout, and Other-voter capture:

$$\mathrm{VSI} = \frac{\text{Votes cast for party i} - \text{Reg. voters for party i}}{\text{Registered voters for Other}}$$

Note the key difference in the numerator. MVE subtracts registered party *i* **ballots cast** (turnout), while VSI subtracts total registered party *i* **voters** (registration). This means the VSI numerator can be decomposed into two components: (1) votes gained from Other-registered voters, minus (2) registered party *i* voters who did not turn out. Larger VSI values are therefore indicative of two factors operating simultaneously: greater turnout among party *i's* registrants, and a greater number of Other-registrant votes flowing to party *i's* candidate.

Because the VSI incorporates differential turnout by using total registered Other voters in the denominator rather than Other ballots cast, the correlation between VSID and VSIR is expected to be weaker than that between MVED and MVER. The data confirm this: in the PA county-level analysis, the MVED-MVER correlation is −0.84, while the VSID-VSIR correlation is −0.71.

In the 2024 Pennsylvania Presidential election, third-party and independent candidates received a combined 67,856 votes — 33,318 for Libertarian candidate Chase Oliver and 34,538 for Green Party candidate Jill Stein — representing approximately 0.97% of total votes cast statewide (Pennsylvania Department of State, 2024). This confirms that Other-registrant ballots flowed overwhelmingly to the two major-party candidates and that the negligible-third-party assumption holds empirically in this context.

Interpreted through the mobilization hypothesis, both indices primarily capture each party's mobilization effectiveness rather than genuine partisan conversion. The MVE approximates net performance relative to the Other-ballot benchmark. The VSI adds a turnout dimension, penalizing parties whose registrants stayed home and rewarding those whose registrants showed up. Together, the two measures provide a more complete picture of each party's electoral performance relative to its registration base.

The following sections apply both indices to the 67 counties of Pennsylvania using data from the 2024 Presidential election.

Data and Descriptive Statistics

The data used in this analysis covers all 67 counties of Pennsylvania. Election data — including voter registration by party affiliation, turnout by party, and presidential vote totals — are drawn from the certified 2024 general election results published by the Pennsylvania Department of State (DoS). All voters not registered as Democrat or Republican are aggregated into the category Other (O), which includes Libertarian, Green, No Affiliation, and all remaining minor party registrations, consistent with PA DoS party code classifications. The four voting indices are computed directly from these county-level registration and turnout figures using the formulas specified in the preceding section; no imputation or estimation is required, as all component values are reported at the county level in the PA DoS files. Socioeconomic and demographic variables are drawn from the U.S. Census Bureau's 2023 five-year American Community Survey (ACS) estimates. We use the 2023 five-year ACS because it was the vintage used when the dataset was constructed, and because five-year ACS estimates provide stable county-level coverage across all 67 Pennsylvania counties. Using this fixed ACS vintage preserves temporal consistency across counties and allows the election indices to be reproduced from a single, clearly specified data release. Counties are matched across datasets on FIPS code. All data sources are publicly available, and the indices can be reproduced directly from the PA DoS election files and ACS summary tables without proprietary data or special access.

Variable Selection

The selection of socioeconomic variables was guided by post-election analysis of the 2024 Presidential results. Exit polling and early demographic breakdowns indicated that income, educational attainment, age, gender composition, population density, and poverty were among the strongest correlates of vote choice (Sanders, 2024; Brownstein, 2025). We supplement these with variables capturing labor force participation, marital status, county land area, and the proportion of foreign-born women aged 15 to 50 — the latter serving as a proxy for immigrant community presence in each county. Counties with larger immigrant populations have historically been targets of intensive Democratic organizing efforts in Pennsylvania, and the variable captures exposure to communities where Democratic mobilization investment has been concentrated; its inclusion therefore reflects both a demographic signal and a theoretically grounded mobilization predictor.

Table 2 presents descriptive statistics for the four index values and 14 socioeconomic variables across all 67 Pennsylvania counties. The counties vary considerably. Median household income ranges from \$47,681 to \$123,041; the share of adults with a bachelor's degree or higher ranges from 11.3% to 56.6%; and population density ranges from 11 to 11,800 persons per square mile. The four indices also exhibit substantial variation. MVED ranges from −1.11 to 1.36, with a mean of 0.29, indicating that, on average across PA counties, Democratic candidates attracted modestly more votes than the number of registered Democratic ballots cast. MVER ranges from −0.02 to 1.94, with a mean of 0.66, reflecting the strong Republican mobilization advantage statewide in 2024. VSID ranges from −1.38 to 0.47, with a mean of −0.15, while VSIR ranges from −0.53 to 0.46, with a mean of −0.06, near zero.

Table 2. Descriptive Statistics for Dependent and Independent Variables (N = 67 PA Counties)

Variable	Description	Mean	Median	SD	Min	Max
Dependent Variables						
MVED	MVE Index, Democrats	0.290	0.367	0.390	-1.111	1.364
MVER	MVE Index, Republicans	0.662	0.579	0.375	-0.017	1.941
VSID	VSI Index, Democrats	-0.154	-0.130	0.351	-1.381	0.466
VSIR	VSI Index, Republicans	-0.057	-0.088	0.234	-0.529	0.459
Economic Variables						
Median HH Income	Median household income ($)	69,614	64,758	13,648	47,681	123,041
Mean HH Income	Mean household income ($)	90,829	83,918	19,178	59,531	165,377
% Below Poverty	% population below poverty level	11.71	11.70	2.900	5.900	22.00
Demographic Variables						
% BA Degree+	% pop. 25+ with bachelor's or higher	25.80	23.30	9.360	11.30	56.60
% Male 20–44	% population male age 20–44	15.38	15.00	2.990	11.10	35.30
Median Age	Median age (years)	43.69	43.70	3.700	33.80	56.10
Sex Ratio	Males per 100 females	102.7	99.30	18.33	90.00	244.4
% Now Married	% pop. 15+ currently married	49.86	50.40	4.860	31.20	56.10
Geographic Variables						
Pop. Density	Population per square mile	481.0	139.0	1488.8	11.00	11,800
Area (sq mi)	County land area (square miles)	667.8	653.2	270.6	130.2	1228.6
Total Pop.	Total county population	193,829	83,213	279,710	4475.0	1,582,432
Labor Market Variables						
LFPR (20–64)	Labor force participation rate, 20–64	75.72	76.40	6.830	34.30	83.90
Emp/Pop Ratio	Employment-population ratio, 20–64	72.07	72.60	6.610	33.20	80.90
Other Variables						
% Women Foreign Born	% women 15–50 foreign born	5.440	3.000	5.430	0.500	22.80

Note. *Data sources: Pennsylvania Department of State (voter registration and election data); U.S. Census Bureau, American Community Survey 2023 5-year estimates (socioeconomic variables). MVE = Measure of Voter Enthusiasm; VSI = Voting Success Index; D = Democrats; R = Republicans.*

Correlation Analysis

Before estimating regression models, we examine bivariate Pearson correlations between all socioeconomic variables and the four indices. Table 3 and Figure 1 present these results. This bivariate examination constitutes a theory-informed variable selection step rather than hypothesis testing; given N = 67 and the resulting constraints on degrees of freedom, it is standard practice in small-N observational research to use bivariate screening to identify candidate predictors before estimating parsimonious multivariate models.

Table 3. Bivariate Pearson Correlations Between Socioeconomic Variables and Voting Indices (N = 67)

Variable	MVED	MVER	VSID	VSIR
Panel A: Socioeconomic Predictors				
Median HH Income	.509***	–.453***	.480***	–.009
Mean HH Income	.513***	–.467***	.449***	.021
% Below Poverty	–.388***	.307**	–.542***	.115
% BA Degree+	.444***	–.412***	.299**	.170
% Male 20–44	.118	–.083	.048	.069
Median Age	–.356***	.271**	–.127	–.014
Sex Ratio (M/F)	–.019	.021	.060	–.037
% Now Married	.108	–.046	.326***	–.259**
Pop. Density	.112	–.123	–.212*	.124
Area (sq mi)	–.056	.199	.022	–.030
Total Pop.	.255**	–.182	–.070	.217*
LFPR (20–64)	.194	–.171	.135	.047
Emp/Pop Ratio	.243**	–.193	.199	.009
% Women Foreign Born	.400***	–.442***	.196	.093
Panel B: Index Intercorrelations				
MVED	—	–.843***	.898***	–.643***
MVER	–.843***	—	–.850***	.671***
VSID	.898***	–.850***	—	–.706***
VSIR	–.643***	.671***	–.706***	—

Note. MVED = Measure of Voter Enthusiasm, Democrats; MVER = Measure of Voter Enthusiasm, Republicans; VSID = Voting Success Index, Democrats; VSIR = Voting Success Index, Republicans. Em dash (—) indicates a diagonal. Leading zeros suppressed per APA convention.

*p < .10. **p < .05. ***p < .01.

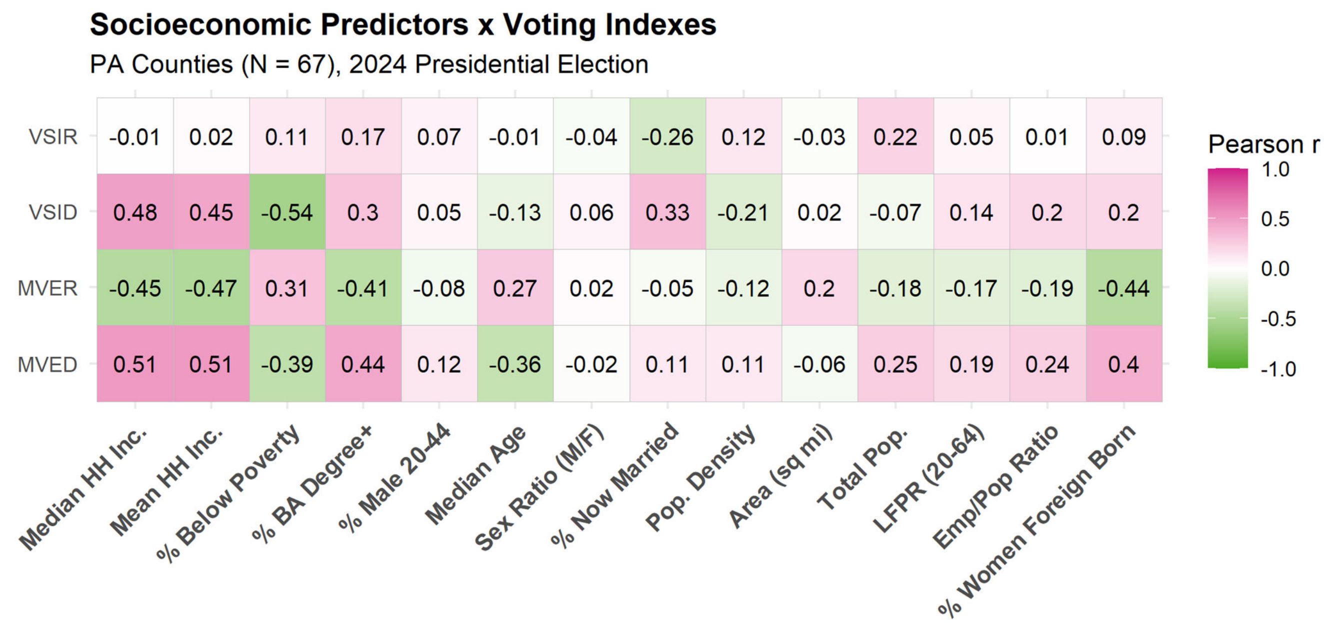


Figure 1. Bivariate Correlations Between Socioeconomic Predictors and Voting Indices Across 67 Pennsylvania Counties, 2024 Presidential Election

Several patterns are immediately apparent. The Democratic indices (MVED, VSID) show strong, consistent correlations with income and economic distress variables. Median household income is positively correlated with both MVED ($r = 0.51$, $p < 0.01$) and VSID ($r = 0.48$, $p < 0.01$), while the poverty rate is negatively correlated with VSID ($r = -0.54$, $p < 0.01$), the single strongest bivariate relationship in the matrix. Educational attainment follows a similar pattern, with the share of adults with a bachelor's degree correlating at $r = 0.44$ with MVED, and $r = 0.30$ with VSID. Median age is negatively associated with MVED ($r = -0.36$, $p < 0.01$), consistent with the generational dimension of the 2024 electorate. The proportion of foreign-born women shows a notable asymmetry, correlating positively with MVED ($r = 0.40$) but negatively with MVER ($r = -0.44$), both significant at $p < 0.01$.

The Republican indices tell a different story. MVER largely mirrors MVED in the opposite direction, as expected given the strong negative correlation between the two ($r = -0.84$). However, VSIR — the Republican Voting Success Index — shows remarkably weak correlations with nearly all socioeconomic variables. Its strongest bivariate correlate is the share of the population currently married ($r = -.26$, $p < .05$), and only one other variable — total population ($r = .22$, $p < .10$) — reaches even marginal significance. This pattern is consistent with the mobilization hypothesis, as VSIR displays comparatively little socioeconomic variation across counties, while Democratic index values are more strongly associated with county-level socioeconomic characteristics.

Collinearity Among Predictors

The correlation matrix also reveals substantial collinearity among the independent variables themselves, which must be addressed before multivariate modeling. Median and mean household income are correlated at $r = .98$. The labor force participation rate and employment-to-population ratio correlate at $r = .99$. Educational attainment correlates with mean household income at $r = .87$ and with median household income at $r = .80$. The poverty rate correlates with median income at $r = -.71$. These collinear pairs cannot appear in the same regression model without inflating standard errors and destabilizing coefficient estimates. The regression specifications described in the following section are designed to avoid these pairings while retaining theoretically motivated predictors from each substantive domain.

DV Intercorrelations

Panel B of Table 3 reports the intercorrelations among the four indices themselves. MVED and MVER correlate at r = −.84, and VSID and VSIR at r = −.71 — both strongly negative, as the index construction implies. MVED and VSID correlate at r = .90, indicating that the two Democratic indices capture substantially overlapping variation despite their different treatment of turnout. The corresponding Republican correlation (MVER-VSIR) is r = .67, somewhat weaker, reflecting the fact that the turnout component introduces more independent variation on the Republican side. Figure 2 illustrates this relationship, highlighting Philadelphia's anomalous position — high MVED but deeply negative VSID — consistent with strong Other-voter capture offset by insufficient Democratic registrant turnout

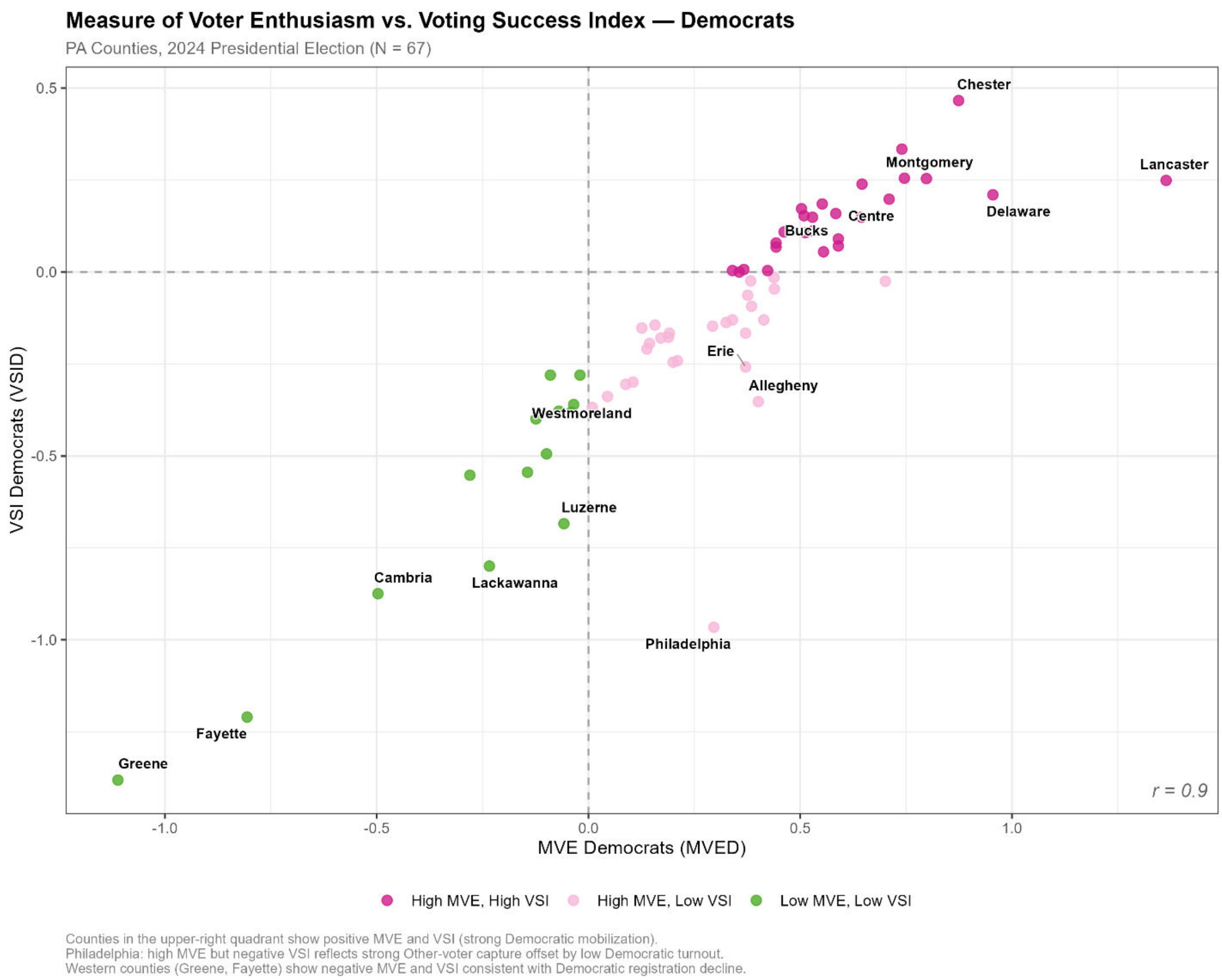


Figure 2. Bivariate Relationship Between MVED and VSID Across 67 Pennsylvania Counties, 2024 Presidential Election

Voting Results, MVE and VSI Descriptive Analysis

Table 4 presents index values and key socioeconomic characteristics for a selection of illustrative counties, spanning the full range of Democratic mobilization outcomes observed across the state. Lancaster County presents the highest MVED in the sample (1.364), reflecting strong Democratic candidate performance among Other-registered voters, consistent with the broader suburban realignment that characterized southeastern Pennsylvania's competitive counties in 2024. Note that because of political registration numbers, even with an MVED greater than one, which implies that some Republican voters voted for the Democratic candidate, the Republican candidate still received more votes than the Democratic candidate.

MVE Indices

The top row of Figure 3 displays the MVE indices. The MVED panel (top left) shows a clear geographic gradient. The southeastern suburban counties — Lancaster, Chester, Montgomery, Delaware, and Bucks — display the warmest tones, indicating that Democratic candidates in these areas attracted substantially more votes than the number of registered Democrats who cast ballots. Philadelphia, despite its large Democratic registration base, shows a moderately positive MVED rather than an extreme value, reflecting the fact that MVE measures performance relative to Other ballots cast rather than absolute vote totals. Centre County, home to Penn State University, also shows a positive MVED, likely reflecting mobilization effects associated with a large university population.

The western and southwestern counties — Greene, Fayette, Washington, Beaver, and Lawrence — display negative MVED values, indicating that the Democratic candidate underperformed relative to registered Democratic turnout in these areas, consistent with either lower Democratic base turnout or Other-registrant support flowing to the Republican party. These are historically Democratic-registration counties that have shifted substantially rightward in recent election cycles, and the negative MVED values quantify the extent of that shift among voters who actually turned out. For example, in Fayette County, in 1952 the Democratic Presidential candidate received 61% of the votes, while in 2024 received 33% of the votes.

The MVER panel (top right) presents a near-perfect mirror image, consistent with the $r = -.84$ intercorrelation between the two indices. Counties with high MVED show low MVER and vice versa. Greene and Fayette display the highest Republican mobilization performance relative to the Other-registrant pool, while the southeastern suburbs and Philadelphia show the lowest. Mirroring is expected from the index construction, as net gains for one party among Other voters must come at the expense of the other party.

Table 4. Voting Indices and Socioeconomic Characteristics for Selected Pennsylvania Counties, 2024 Presidential Election

	MVED	MVER	VSID	VSIR	Median HH Income	% Poverty	% BA+	% Foreign Born Women	Median Age	Pop. Density
Lancaster	1.364	0.975	0.249	-0.214	$83,703	8.7	31.0	8.5	39.1	588
Chester	0.874	0.067	0.466	-0.157	$123,041	5.9	56.6	15.2	40.7	721
Montgomery	0.746	0.191	0.255	-0.064	$111,521	6.5	52.3	16.1	41.1	1,783
Centre	0.710	0.251	0.198	-0.088	$72,748	16.9	46.5	12.9	33.8	142
Allegheny	0.401	0.666	-0.352	0.196	$76,393	11.2	44.8	8.9	40.6	1,699
Erie	0.371	0.579	-0.258	0.008	$61,476	15.4	29.6	5.8	40.1	337
Philadelphia	0.296	0.620	-0.966	0.142	$60,698	22.0	34.6	18.5	35.1	11,800
Luzerne	-0.058	1.119	-0.685	0.161	$62,321	15.4	25.0	15.1	42.1	366
Fayette	-0.806	1.779	-1.210	0.459	$56,093	17.4	18.9	1.1	45.4	161
Greene	-1.111	1.941	-1.381	0.367	$66,870	12.5	21.1	0.9	43.0	61

Note. Selected counties representing high Democratic performance (Chester, Montgomery, Lancaster), large urban centers (Philadelphia, Allegheny), swing counties (Erie, Luzerne), and historically Democratic counties with strong Republican shift (Fayette, Greene).
MVED = Measure of Voter Enthusiasm, Democrats; MVER = Measure of Voter Enthusiasm, Republicans; VSID = Voting Success Index, Democrats; VSIR = Voting Success Index, Republicans. % BA+ = percentage of adults 25+ with bachelor's degree or higher. % Foreign Born Women = percentage of women aged 15–50 who are foreign born. Pop. Density = persons per square mile.
Socioeconomic data from U.S. Census Bureau, American Community Survey 2023 5-year estimates.

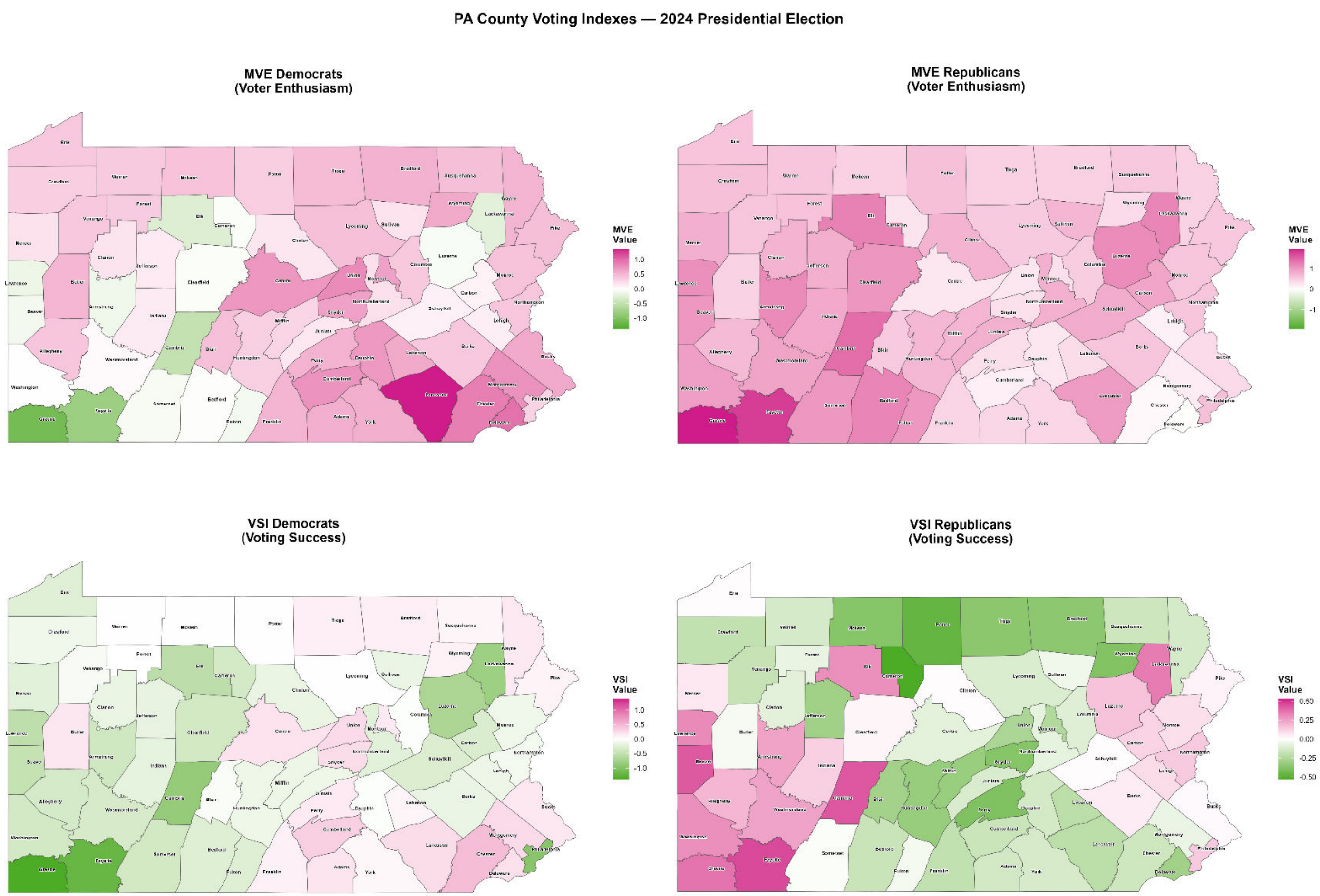


Figure 3. County-Level Voting Indices Across 67 Pennsylvania Counties, 2024 Presidential Election. Top row: Measure of Voter Enthusiasm (MVE) for Democrats (left) and Republicans (right). Bottom row: Voting Success Index (VSI) for Democrats (left) and Republicans (right). Pink/magenta indicates positive index values; green indicates negative values. Note the compressed scale on the VSIR panel (bottom right), reflecting the narrow range and lack of spatial variation in Republican voting success.

VSI Indices

The bottom row displays the VSI indices, which incorporate turnout and Other-voter capture. The VSID panel (bottom left) shows a spatial pattern similar to MVED but with some notable differences. The southeastern suburban cluster remains positive, and the western counties remain negative. However, the overall distribution is shifted downward — the mean VSID is −.15, compared to a mean MVED of .29 — reflecting the fact that VSI penalizes parties whose registrants fail to turn out. Counties with high Democratic registration but lagging Democratic turnout show lower VSID values than their MVED values would suggest. Philadelphia is a clear example. Its MVED is moderately positive, but its VSID is negative, indicating that despite strong Other-vote capture, insufficient Democratic turnout dragged down overall Democratic voting success.

The VSIR panel (bottom right) is the most striking of the four. Unlike the other three panels, which display clear spatial gradients, VSIR shows no discernible geographic pattern. The color scale is compressed (ranging only from approximately −.53 to .46, compared to the wider ranges on the other panels), and the distribution of colors across the map appears essentially random with respect to the urban-rural, east-west, and income gradients visible in the other three panels. This visual impression confirms the statistical finding from the correlation analysis. Republican VSIR values were not concentrated in any particular county type and displayed comparatively little observable geographic or socioeconomic clustering. This is consistent with the mobilization hypothesis. At the county level, Republican voting success showed comparatively little association with socioeconomic structure, whereas Democratic index values were more strongly associated with county-level socioeconomic characteristics.

Methodology and Results

The analytic approach proceeds in two stages. The first, presented in the preceding sections, examines bivariate correlations between the four voting indices and 14 socioeconomic variables to identify the strongest and most theoretically relevant predictors. The second, presented here, estimates parsimonious OLS regression models to confirm that these bivariate patterns hold in a multivariate context.

Model Specifications

Using *i* to index counties and *p* for party, the basic specification is:

$$\widehat{Index}_{i,p} = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_{1i} + \hat{\beta}_2 \cdot X_{2i} + \cdots + \hat{\beta}_k \cdot X_{ki}$$

where the dependent variable is one of the four indices (MVED, MVER, VSID, or VSIR) and the independent variables are county-level socioeconomic characteristics drawn from the Census ACS.

With only 67 observations, model parsimony is essential to avoid overfitting. We limit each specification to three or four predictors — well within the conventional guideline of at least 15 observations per predictor (Stevens, 1996). Variable selection is guided by three criteria: (a) the strength of bivariate correlations identified in the preceding section, (b) theoretical relevance to the 2024 election as informed by post-election analysis (Sanders, 2024; Brownstein, 2025), and (c) avoidance of collinear predictor pairs. As noted earlier, several pairs of predictors exhibit correlations above $|r| = .70$, including median and mean household income ($r = .98$), the labor force participation rate and employment-population ratio ($r = .99$), and educational attainment and household income ($r = .80$–$.87$). These pairs are never included in the same model.

All models are estimated using ordinary least squares with HC3 heteroskedasticity-consistent standard errors (MacKinnon & White, 1985), which provide reliable inference under heteroskedasticity without requiring correct specification of the error variance structure. HC3 standard errors are conservative relative to classical standard errors, particularly in small samples, making them the appropriate choice given $N = 67$. Variance inflation factors were computed for each specification to confirm that multicollinearity is within acceptable bounds. All VIFs fall below 2.25 across both models, well under conventional thresholds (see table notes). The Core SES Model (Table 5) includes median household income, the proportion of foreign-born women aged 15 to 50, and median age — three predictors representing economic standing, immigration and diversity, and the generational divide, respectively. All VIFs are below 2.2. The Best Correlates Model (Table 6) includes the poverty rate, the proportion of foreign-born women, median age, and the share of the population currently married — four predictors that draw from different substantive domains while avoiding collinear pairs. All VIFs are below 2.25.

Results: Core SES Model (Table 5)

The Core SES Model explains a meaningful share of variance in the three Democratic-side indices. Adjusted R^2 values are .27 for MVED, .21 for MVER, and .22 for VSID. The overall F-statistics are significant at $p < .01$ for all three. For VSIR, by contrast, the model explains no variance (adjusted $R^2 = -.03$; $F = 0.59$, n.s.).

Table 5. OLS Regression Results — Core SES Model: Median Household Income, Proportion Foreign-Born Women, and Median Age as Predictors of County-Level Voting Indices (HC3 Robust Standard Errors). N = 67 Pennsylvania counties. All VIFs < 2.2.

Parsimonious OLS — Core SES (Robust HC3 SE)

	Dependent variable:			
	MVED	MVER	VSID	VSIR
	(1)	(2)	(3)	(4)
Median HH Income (log)	0.9811***	-0.6047*	1.1907***	-0.1208
	(0.2743)	(0.3249)	(0.3270)	(0.2026)
% Women Foreign Born (15-50)	0.00118	-0.0167	-0.0124	0.00768
	(0.0129)	(0.0129)	(0.0125)	(0.00704)
Median Age	-0.0198	0.00369	-0.00168	0.00322
	(0.0121)	(0.0137)	(0.0115)	(0.00877)
% Now Married (15+)				
Population Density				
Constant	-9.774***	7.324**	-13.270***	1.105
	(2.788)	(3.253)	(3.623)	(2.073)
Observations	67	67	67	67
R^2	0.304	0.244	0.253	0.015
Adjusted R^2	0.271	0.208	0.217	-0.031
Residual Std. Error (df = 63)	0.333	0.334	0.310	0.237
F Statistic (df = 3; 63)	9.192***	6.793***	7.109***	0.330
Note:	*p<0.1; **p<0.05; ***p<0.01			

Median household income is the dominant predictor. A one percent increase in median county income is associated with an increase of approximately 0.0098 in MVED ($p < .01$), a decrease of roughly 0.0060 in MVER ($p < .10$), and an increase of approximately 0.0119 in VSID ($p < .01$). To contextualize these magnitudes: moving from the lowest-income county to the highest represents a difference of approximately 0.95 log units (ln($123,041) − ln($47,681)), which corresponds to a predicted increase in MVED of roughly 0.93 and in VSID of roughly 1.13 — spanning most of the observed range of both indices. The effect on VSIR is negligible and statistically insignificant.

Median age does not reach significance in any equation in this specification, though the negative coefficient for MVED ($b = -.020$) is consistent with younger counties showing stronger Democratic mobilization performance. The proportion of foreign-born women similarly fails to reach significance in any equation, suggesting that its bivariate correlation with the indices ($r = .40$ with MVED, $r = -.44$ with MVER) is substantially mediated by income.

Results: Best Correlates Model (Table 6)

The Best Correlates Model achieves somewhat higher explanatory power: adjusted R^2 values of .29 for MVED, .22 for MVER, and .29 for VSID. All three F statistics are significant at $p < .01$. VSIR again shows no meaningful fit (adjusted $R^2 = .03$; $F = 0.82$, n.s.).

The poverty rate is the strongest predictor in this specification. Each percentage-point increase in county poverty is associated with a decrease of .053 in MVED ($p < .05$) and .062 in VSID ($p < .05$). These effects are robust to the HC3 correction. Poverty does not significantly predict MVER or VSIR. The proportion of foreign-born women, which was insignificant in the Core SES Model, emerges as a significant predictor of MVER in this specification ($b = -.023$, $p < .05$), suggesting that counties with larger immigrant communities display lower Republican mobilization success even after controlling for poverty, age, and marital status. Median age is again marginally significant for MVED ($b = -.031$, $p < .10$). The share of the population currently married does not reach significance in any equation.

Table 6. OLS Regression Results — Best Correlates Model: Poverty Rate, Proportion Foreign-Born Women, Median Age, and Marital Status as Predictors of County-Level Voting Indices (HC3 Robust Standard Errors)
N = 67 Pennsylvania counties. All VIFs < 2.25.

Parsimonious OLS — Best Correlates (Mixed) (Robust HC3 SE)

	Dependent variable:			
	MVED	MVER	VSID	VSIR
	(1)	(2)	(3)	(4)
% Below Poverty Level	-0.0528**	0.0377	-0.0621**	-0.00749
	(0.0246)	(0.0249)	(0.0276)	(0.0209)
% Women Foreign Born (15-50)	0.0133	-0.0233**	0.00276	0.00347
	(0.0109)	(0.00909)	(0.00774)	(0.00592)
Median Age	-0.0306*	0.0103	-0.0165	0.00932
	(0.0173)	(0.0170)	(0.0175)	(0.0130)
% Now Married (15+)	-0.00033	0.00249	0.00441	-0.0170
	(0.0204)	(0.0138)	(0.0245)	(0.0173)
Population Density				
Constant	2.191**	-0.229	1.061	0.454
	(1.084)	(1.151)	(1.093)	(0.964)
Observations	67	67	67	67
R^2	0.329	0.265	0.329	0.087
Adjusted R^2	0.286	0.217	0.285	0.028
Residual Std. Error (df = 62)	0.330	0.332	0.297	0.230
F Statistic (df = 4; 62)	7.609***	5.579***	7.592***	1.469
Note:		*p<0.1; **p<0.05; ***p<0.01		

The VSIR Null Findings

Across both full-sample specifications, the included county-level socioeconomic predictors explain little variation in the Republican Voting Success Index. Neither model achieves a significant F-statistic for the VSIR equation. No individual predictor reaches conventional levels of significance. The adjusted R^2 values are negative or near zero. This pattern stands in stark contrast to the Democratic indices, where income, poverty, and other SES variables explain 20 to 29 percent of the variance.

This asymmetry is the paper's central empirical finding. Interpreted through the mobilization framework, it indicates that Republican voting success displayed comparatively little socioeconomic structuring across counties. Republican voting success showed no meaningful county-level variation along wealth, education, or urban-rural dimensions. Democratic voting success, by contrast, was strongly conditioned by county-level socioeconomic characteristics, with Democratic indices performing better in higher-income, lower-poverty, younger counties. The practical implication is that Democratic campaigns face a more geographically differentiated mobilization challenge than Republican campaigns do, at least in the 2024 Pennsylvania context. Whether this uniformity reflects a durable structural feature of Republican mobilization or a condition specific to 2024 — when Republicans constituted the surge party in the sense of Campbell (1960) & J. E. Campbell (1987) — is a question the indices are well-positioned to answer as future electoral data become available.

Influence Diagnostics and Robustness

To assess whether results are sensitive to individual influential observations, Cook's distance was computed for all models using the conventional threshold of $4/N = 0.060$, and leave-one-out (LOO) analyses were conducted excluding each flagged county in turn. Nine counties exceeded the threshold in at least one model: Forest, Philadelphia, Cameron, Greene, Fayette, Lehigh, Centre, Luzerne, & Lancaster (see Appendix A).

The Core SES Model findings are robust across all exclusions. The income coefficient on MVED remains positive and significant at $p < .01$ under every county exclusion, with coefficients ranging from 0.886 to 1.102 — a narrow band that confirms the income-mobilization gradient is not an artifact of any single county. The VSIR null finding is equally stable: the income coefficient on VSIR never approaches conventional significance thresholds under any exclusion scenario, with the largest t-statistic observed when Cameron is excluded ($p = .142$). The VSIR null finding requires one additional note. When Forest County is excluded from the Best Correlates Model, the overall model F reaches marginal significance ($p = .039$, adj. $R^2 = .095$); however, the poverty coefficient in that specification remains substantively negligible and non-significant ($b = -0.021$, $p = .253$). The marginal model F reflects modest joint predictive power from the other covariates — median age, marital status, and the proportion of foreign-born women — not a poverty effect on Republican mobilization. The Core SES Model VSIR equation, which uses income as the primary predictor, remains non-significant across all exclusion scenarios (p range: .553–.904). The VSIR null finding is robust.

The effects of poverty in the Best Correlates Model warrant additional discussion. The poverty coefficient on VSID remains statistically significant ($p < .05$) across all exclusions, including Forest ($b = -0.043$, $p = .041$), providing consistent support for that finding. The poverty effect on MVED is directionally consistent across all exclusions — the coefficient is negative in every specification — but loses conventional significance when Forest is excluded ($b = -0.038$, $p = .102$). Forest County (population approximately 7,200) has an unusually small Other-registrant denominator, resulting in index values with high leverage in the Best Correlates model specifications. The SES Model income result, which is stable across all exclusions, provides a more reliable estimate of the socioeconomic gradient for Democratic mobilization. Given the inherent constraints of a 67-county sample, the directional consistency of the poverty effect across all exclusion scenarios is itself a meaningful robustness signal — one that replication across additional election cycles would sharpen. Full diagnostic results are reported in Appendix A.

Conclusion

This paper introduces two new indices — the Measure of Voter Enthusiasm (MVE) and the Voting Success Index (VSI) — designed to quantify a party's electoral performance relative to its registration base at the sub-state level. The MVE approximates net vote performance relative to the Other-ballot benchmark, or more precisely, the net votes a party's candidate attracts beyond its own registrants' turnout, normalized against the Other-registrant pool. The VSI extends this by penalizing parties whose registrants stay home and rewarding those whose registrants show up. Applied to Pennsylvania's 67 counties using data from the 2024 Presidential election, the indices reveal a clear and consequential asymmetry in the socioeconomic structure of partisan mobilization.

Democratic index values — MVED and VSID — are strongly correlated with county-level income, poverty, educational attainment, median age, and the proportion of foreign-born residents. Parsimonious OLS models explain 22-29% of the variance in these indices using three or four socioeconomic predictors. Higher-income, lower-poverty, more educated, and younger counties show higher Democratic index values, while economically distressed counties show weaker Democratic performance. These patterns are visible both statistically, in the correlation and regression analyses, and spatially, in the choropleth maps that reveal a clear geographic gradient from the southeastern suburbs to the western and rural interior.

Republican index values tell a fundamentally different story. MVER largely mirrors MVED in the opposite direction, as the index construction implies. But VSIR — the Republican Voting Success Index — is essentially uncorrelated with standard socioeconomic variables. No full-sample regression specification yields a significant F statistic for VSIR, and no individual predictor is significant. Republican voting success in Pennsylvania for 2024 displayed no meaningful concentration by county type; VSIR showed comparably uniform values across the state. This finding is consistent with the stability of partisan identification documented in the political behavior literature (Green, Palmquist, & Schickler, 2002; Keith et al., 1992). The indices capture differential mobilization and independent capture, not partisan conversion, and the Republican county-level voting success in 2024 displayed comparatively little socioeconomic variation, unlike Democratic county-level index values.

Several limitations should be acknowledged. First, with $N = 67$, statistical power is limited, and the models are intentionally parsimonious. County-level analysis may mask important within-county variation; municipality-level or precinct-level data, if available statewide, would substantially increase both power and geographic resolution. Second, the analysis is cross-sectional, capturing a single election. Whether the patterns observed here, particularly the VSIR null findings, are specific to the 2024 contest or reflect a durable structural feature of Pennsylvania politics is an empirical question that can be addressed only through temporal comparisons. Third, the ecological fallacy remains a concern. County-level correlations between socioeconomic characteristics and voting indices do not establish that individual voters with particular characteristics behaved in particular ways. The indices describe aggregate mobilization outcomes, not individual vote choices. The absence of significant VSIR relationships may reflect relative uniformity, measurement limitations, restricted variance, omitted variables, or insufficient statistical power given $N = 67$.

The indices cannot perfectly separate the mobilization and independent-capture mechanisms they are designed to measure from other potential sources of index variation, including differential ballot roll-off, localized candidate effects, or modest amounts of partisan crossover. However, under conditions of high partisan polarization, the primary source of index variation is differential base turnout — whether a party's registered voters show up at rates consistent with, above, or below the competitive benchmark — which is the core prediction of surge and decline theory (Campbell, 1960) and consistent with the small third-party vote share documented above. We present the mobilization interpretation as the most theoretically coherent account of index variation rather than a definitive inference.

These limitations suggest several directions for future research. Extending the analysis to municipality-level data across all of Pennsylvania would provide a much richer picture of mobilization geography. Applying the indices to prior elections (2016, 2020) would reveal whether the Democratic SES gradient and the comparatively low socioeconomic variation in VSIR are stable features or specific to 2024. The indices are designed to be applicable in any closed-primary state where voter registration by party affiliation is publicly available, and comparative analysis across states — particularly other 2024 battlegrounds such as Michigan, Wisconsin, and Arizona — would test the generalizability of the mobilization asymmetry documented here. Finally, the practical implications for campaign strategy deserve direct attention. If Democratic mobilization success is strongly conditioned by local socioeconomic context, while Republican success is not, the two parties face structurally different resource-allocation problems, and the indices provide a quantitative framework for identifying where marginal campaign investments are most likely to shift outcomes.

More broadly, the surge-and-decline theory (A. Campbell, 1960; J. E. Campbell, 1987) offers a compelling prediction for future electoral cycles. Whichever party occupies the out-party position should exhibit more uniform sub-state mobilization, while the declining party's index values should become increasingly structured by local socioeconomic context. If the asymmetry documented here reverses in 2028 — with Republican VSI values beginning to mirror the SES-structured fragmentation currently observed in Democratic indices — that would provide strong evidence that the VSIR null finding reflects surge conditions rather than a durable structural feature of Republican mobilization.

APPENDIX A: Influence Diagnostics and Leave-One-Out Robustness

Cook's distance was computed for all models using the conventional threshold of $4/N = 0.060$. Table A1 identifies all counties flagged above this threshold. Table A2 reports leave-one-out adjusted R^2 values for each flagged county across the five outcome equations most relevant to the paper's central findings. Table A3 reports HC3 robust coefficient estimates for the key predictors under each exclusion.

Table A1. Counties Flagged by Cook's Distance (threshold = 4/67 = 0.060)

County	Models Flagged	Max Cook's *D*	Note
Forest	6	0.983	Pop. approx. 7,200; small Other-registrant denominator
Philadelphia	2	0.311	Anomalous registration-to-turnout gap; discussed in the main text
Cameron	4	0.261	Pop. approx. 4,500; small Other-registrant denominator
Greene	6	0.180	Historically Democratic County with pronounced 2024 Republican shift
Lehigh	2	0.153	High Democratic Other-capture
Centre	1	0.145	Atypical age and education profile
Luzerne	2	0.135	Swing county; discussed in the main text
Fayette	7	0.127	Historically Democratic County with pronounced 2024 Republican shift
Lancaster	2	0.095	High Democratic Other-capture

Note. Models flagged = number of models (of 8 total) in which Cook's *D* exceeds the threshold. Forest & Cameron leverage reflects small-denominator instability rather than substantive extremity.

Table A2. Leave-One-Out Robustness: Adjusted R^2 by County Exclusion

Excluded County	Core SES MVED	Core SES VSID	Core SES VSIR	Best Correlates VSID	Best Correlates VSIR
Full sample (N = 67)	.271	.217	−.031	.285	.028
Forest (N = 66)	.297***	.253***	−.031	.347***	.095*
Philadelphia (N = 66)	.271***	.197***	−.039	.237***	.016
Cameron (N = 66)	.268***	.217***	−.007	.289***	.058
Greene (N = 66)	.350***	.296***	−.020	.355***	.049
Fayette (N = 66)	.261***	.207***	−.028	.240***	.043
Lehigh (N = 66)	.290***	.217***	−.038	.286***	.024
Centre (N = 66)	.261***	.224***	−.032	.318***	.028
Luzerne (N = 66)	.273***	.202***	−.040	.275***	.019
Lancaster (N = 66)	.251***	.206***	−.033	.272***	.021

Note. *** $p < .001$; * $p < .05$. Significance indicators reflect model-level *F* p-values. Core SES VSIR and Best Correlates VSIR remain non-significant across all exclusions except Forest, where Best Correlates VSIR reaches $p = .039$; however, the poverty coefficient in that specification is non-significant ($b = -0.021$, $p = .253$), confirming the marginal *F* reflects other predictors rather than a substantive poverty effect on Republican mobilization.

Table A3. HC3 Robust Coefficient Stability: Key Predictors Across County Exclusions

Sample	**log(Income) → MVED**	**log(Income) → VSID**	**log(Income) → VSIR**	**Poverty → VSID**
Full (N = 67)	0.981***	1.191***	−0.121	−0.062**
Excl. Forest	1.102***	1.324***	−0.135	−0.043**
Excl. Philadelphia	0.955***	0.969***	−0.073	−0.055**
Excl. Cameron	1.033***	1.235***	−0.246	−0.066**
Excl. Greene	1.061***	1.263***	−0.147	−0.059**
Excl. Fayette	0.901***	1.110***	−0.072	−0.051†
Excl. Lehigh	0.886***	1.206***	−0.110	−0.061**
Excl. Centre	0.996***	1.224***	−0.125	−0.068**
Excl. Luzerne	0.899***	1.108***	−0.083	−0.058**
Excl. Lancaster	0.890***	1.168***	−0.104	−0.062**

Note. *** $p < .001$; ** $p < .05$; † $p < .10$. All standard errors HC3 heteroskedasticity-consistent. log(Income) = log-transformed median household income; Poverty = percent population below poverty level. Both predictors from their respective model specifications: income from the Core SES Model, poverty from the Best Correlates Model. The income effect on MVED and VSID is significant at $p < .001$ across all exclusions. The VSIR null finding is stable throughout. The poverty effect on VSID is significant at $p < .05$ in all but one exclusion (Fayette, $p = .059$). The poverty effect on MVED is directionally consistent but sensitive to Forest inclusion; see main text discussion.

References

American National Election Studies. (2025). ANES 2024 Time Series Study full release [Dataset and documentation]. August 8, 2025 version. www.electionstudies.org

Bartels, L. M. (2000). Partisanship and voting behavior, 1952–1996. *American Journal of Political Science, 44*(1), 35–50. https://doi.org/10.2307/2669291

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. Political Behavior, 24(2), 117–150. https://doi.org/10.1023/A:1021226224601

Berinsky, A. J. (2005). The perverse consequences of electoral reform in the United States. American Politics Research, 33(4), 471–491. https://doi.org/10.1177/1532673X042694

Biggers, D. R., & Hanmer, M. J. (2017). Understanding the adoption of voter identification laws in the American States. American Politics Research, 45(4), 560–588. https://doi.org/10.1177/1532673X16687266

Brownstein, R. (2025, May 11). Trump made historic gains with minority voters in 2024. They are already pulling back in 2025. CNN. https://www.cnn.com/2025/05/11/politics/gop-voter-coalition-2024-trump

Campbell, A. (1960). Surge and decline: A study of electoral change. Public Opinion Quarterly, 24(3), 397–418. http://www.jstor.org/stable/2746724

Campbell, J. E. (1987). The revised theory of surge and decline. American Journal of Political Science, 31(4), 965– 979. https://doi.org/10.2307/2111231

Campbell, J. E. (2008). The trial-heat forecast of the 2008 presidential vote: Performance and value considerations in an open-seat election. PS: Political Science & Politics, 41(4), 697–701. https://doi.org/10.1017/S1049096508081158

Cook Political Report. (n.d.). The Cook Partisan Voting Index (Cook PVI). https://www.cookpolitical.com/cook-pvi

Dalton, R. J. (2008). *The good citizen: How a younger generation is reshaping American politics*. CQ Press.

Elliott, K. (2023). What is it like to be a partisan? Measures of partisanship and its value for democracy. Perspectives on Politics, 22(1), 1–15. https://doi.org/10.1017/S153759272300289X

Evans, G., & Andersen, R. (2006). The political conditioning of economic perceptions. The Journal of Politics, 68(1), 194–207. https://doi.org/10.1111/j.1468-2508.2006.00380.x

Fraga, B. L. (2018). The turnout gap: Race, ethnicity, and political inequality in a diversifying America. Cambridge University Press.

Gallagher, M. (1991). Proportionality, disproportionality, and electoral systems. Electoral Studies, 10(1), 33–51. https://doi.org/10.1016/0261-3794(91)90004-C

Gimpel, J. G., & Schuknecht, J. E. (2003). Political participation and the accessibility of the ballot box. Political Geography, 22(5), 471–488. https://doi.org/10.1016/S0962-6298(03)00029-5

Golosov, G. V. (2024). Electoral rules and the two-party system: A methodological inquiry with reference to Duverger's law. International Political Science Review, 46(3), 388-405. https://doi.org/10.1177/01925121241283715

Green, D., Palmquist, B., & Schickler, E. (2002). Partisan hearts and minds: Political parties and the social identities of voters. Yale University Press.

Green, J., & McElwee, S. (2019). The differential effects of economic conditions and racial attitudes in the election of Donald Trump. Perspectives on Politics, 17(2), 358–379. https://doi.org/10.1017/S1537592718003365

Kalmoe, N. P., & Mason, L. (2022). Radical American partisanship: Mapping violent hostility, its causes, and the consequences for democracy. University of Chicago Press.

Kavanagh, J., & Rich, M. D. (2018). Truth decay: An initial exploration of the diminishing role of facts and analysis in American public life. RAND Corporation. https://www.rand.org/pubs/research_reports/RR2314.html

Keith, B. E., Magleby, D. B., Nelson, C. J., Orr, E., Westlye, M. C., & Wolfinger, R. E. (1992). The myth of the independent voter. University of California Press.

Key, V. O. (1949). Southern politics in state and nation. Knopf.

Levendusky, M. S., Pope, J. C., & Jackman, S. D. (2008). Measuring district-level partisanship with implications for the analysis of U.S. elections. The Journal of Politics, 70(3), 736–753. https://doi.org/10.1017/S0022381608080729

Lewis-Beck, C., & Martini, N. F. (2020). Economic perceptions and voting behavior in US presidential elections. Research & Politics, 7(4), 1–6. https://doi.org/10.1177/2053168020972811

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. Journal of Econometrics, 29(3), 305–325. https://doi.org/10.1016/0304-4076(85)90158-7

Mason, L. (2018). Uncivil agreement: How politics became our identity. University of Chicago Press.

McCarty, N., Poole, K. T., & Rosenthal, H. (2006). Polarized America: The dance of ideology and unequal riches. MIT Press.

Paul, M., Zhang, R., Liu, B., et al. (2022). State-level political partisanship strongly correlates with health outcomes for US children. European Journal of Pediatrics, 181, 273–280. https://doi.org/10.1007/s00431-021-04203-y

Pennsylvania Department of State. (2024). Governor Shapiro certifies 2024 presidential election results. Commonwealth of Pennsylvania. https://www.pa.gov/agencies/dos/newsroom/governor-shapiro-certifies-2024-presidential-election-results

Petrocik, J. R. (1981). Party coalitions: Realignments and the decline of the New Deal party system. University of Chicago Press.

Poole, K. T., & Rosenthal, H. (1985). A spatial model for legislative roll call analysis. American Journal of Political Science, 29(2), 357–384. https://doi.org/10.2307/2111172

Rosenstone, S. J., & Hansen, J. M. (1993). Mobilization, participation, and democracy in America. Macmillan.

Sanders, L. (2024, November 7). How 5 key demographic groups voted in 2024: AP VoteCast. Associated Press. https://apnews.com/article/election-harris-trump-women-latinos-black-voters-0f3fbda3362f3dcfe41aa6b858f22d12

Shaw, D. (2012). If everyone votes their party, why do presidential election outcomes vary so much? The Forum, 10(3). https://doi.org/10.1515/1540-8884.1519

Sides, J., Tesler, M., & Vavreck, L. (2018). Identity crisis: The 2016 presidential campaign and the battle for the meaning of America. Princeton University Press.

Skocpol, T., & Fiorina, M. P. (Eds.). (1999). Civic engagement in American democracy. Brookings Institution Press.

Stevens, J. P. (1996). Applied multivariate statistics for the social sciences (3rd ed.). Lawrence Erlbaum Associates.

Subramanian, S. V., Jones, K., Kaddour, A., & Krieger, N. (2009). Revisiting Robinson: The perils of individualistic and ecologic fallacy. International Journal of Epidemiology, 38(2), 342–360. https://doi.org/10.1093/ije/dyn359

Tesler, M. (2016). Post-racial or most-racial? Race and politics in the Obama era. University of Chicago Press.