ISSN: 3049-7159

Review Article

Journal of Business and Econometrics Studies

Measurement Error Diagnostics in Digital Data

Richard Mulenga

Department of Economics, ZCAS University, Lusaka, Zambia

Corresponding author
Richard Mulenga, Department of Economics, ZCAS University, Lusaka, Zambia.

Received: November 10, 2025; **Accepted:** November 18, 2025; **Published:** November 25, 2025

ABSTRACT
This study examines the use of the reverse regression (RR) model as a diagnostic instrument for detecting and correcting bias in the digital datasets by employing digitally simulated data from a hypothetical medium-sized online enterprise. Set against the backdrop of the high-frequency data generation by digital platforms, this study tackles the widespread issue of measurement error and endogeneity that arises from self-reported, algorithmically modified, and poorly validated datasets. The primary objective was to assess the effectiveness of RR in detecting bias when the assumptions of ordinary least squares (OLS) are not met, and to evaluate its potential for enhancing the reliability of empirical analysis in datasets derived from various platforms. The results indicate that RR enhances model robustness by identifying distortions in parameter estimates and facilitates more precise causal interpretations when used in conjunction with instrumental variables (IVs) and the generalised method of moments (GMM) framework. Empirical findings show that, ceteris paribus, a one-unit increase in actual digital advertising expenditure leads to a 5.04% rise in digital sales, whereas a similar increase in reported advertising expenditure results in a 4.78% increase in sales. In the reverse regressions, both actual and reported advertising expenditures remain positive and statistically significant, with coefficients of 3.06% and 2.72%, respectively. These consistent findings confirm that investments in advertising have a significant impact on online sales performance and consumer engagement. The study concludes by suggesting a policy-oriented framework aimed at enhancing data reliability and informing the formulation of digital economy policies, particularly in calibrating tax incentives, innovation grants, and digital interventions based on empirically validated metrics.

JEL Classifications: C18, L86, 033

Keywords: Digital Platforms; Measurement Error, Reverse Regressions, Simulation Analytics

Introduction and Motivation

This research evaluates the use of reverse regression as a diagnostic tool set for detecting bias within datasets of the digital economy, integrating both simulated scenarios and real-world applications. The study is based on the premise that digital platforms consistently produce vast datasets, which are used by researchers and policymakers to examine market dynamics, consumer behaviour, and employment trends. However, these datasets frequently suffer from measurement errors arising from self-reported data, algorithmic modifications, and inadequate validation procedures. Additionally, endogeneity problems stemming from simultaneity and omitted variable bias (OVB) further undermine causal inference in the analysis of digital data.

Reverse regressions (RRs) help identify and assess bias when Ordinary Least Squares (OLS) assumptions are violated. This method is particularly useful for online advertising attribution, self-reported gig earnings, and behavioural data from digital platforms. Employing simulated data from a mid-sized company's digital sales and advertising from platforms like Google Ads and Analytics Advertising Dashboards (AADs) . The digital economy has changed how data is produced, gathered, and analysed. Data from digital platforms, ranging from e-commerce transactions to user-generated content, offer vast potential for economic analysis; however, they frequently suffer from measurement errors and endogeneity problems. The proliferation of digital platforms, such as Uber and Airbnb, has transformed digital markets, consumer habits, and economic frameworks. Digital platforms enable users to connect, communicate, and engage with one another. Furthermore, digital platforms provide the necessary infrastructure for the buying, selling, and exchanging of goods and services [1,2]. Reverse regressions (RRs), initially developed to detect measurement errors, appear not to be widely

Citation: Richard Mulenga. Measurement Error Diagnostics in Digital Data. J Bus Econo Stud. 2025. 2(6): 1-12. DOI: doi.org/10.61440/JBES.2025.v2.96

used in the expanding research on digital economy platform datasets, despite their significant potential for mitigating empirical digital data bias and endogeneity [3,4]. Therefore, this study aims to address this gap by evaluating the use of RR to identify and quantify measurement errors (MEs) and endogeneity within datasets derived from digital platforms. The research utilises simulated data from digital platform datasets to illustrate the practical applications of RR techniques as tools for detecting empirical bias in digital data. Datasets obtained from various digital platforms, including Uber, Airbnb, Facebook, Amazon, Spotify, and Google, offer valuable insights into user behaviour and market dynamics. However, it is imperative to recognise that these datasets were not generated initially with the goal of facilitating academic research. This creates specific challenges for researchers striving to utilise this data effectively [5].

Measurement errors (MEs) can originate from numerous sources, including the design characteristics of platforms such as rating systems, user behaviours like self-selection, and data processing methods, including search ranking algorithms. Moreover, endogeneity issues frequently arise in data obtained from digital platforms, which can be attributed to the dynamics of two-sided markets and the existence of feedback loops among various variables. Data sourced from digital platforms often includes variables that are prone to noise, leading to widespread measurement errors. For example, ratings may be influenced by subjective biases and social desirability effects ([6,7]. Location data may lack accuracy, as it typically relies on IP-based geolocation techniques. Additionally, prices often undergo dynamic and algorithmic adjustments, creating uncertainty regarding their true observed values [8]. These digital data errors undermine the reliability of critical regressors and the availability of strong instrumental variables, which may result in biased coefficient estimates in empirical forward (direct) regression analyses [9]. In digital marketplaces, simultaneity or endogeneity issues are commonly encountered, where prices and demand levels mutually influence one another in real time [10]. Furthermore, platform visibility elements, such as search rankings, are affected and shaped by click-through rates [11]. Additionally, unobserved factors such as alterations in platform policies or consumer expectations can concurrently impact both the explanatory variables and the outcomes being analysed.

Recent advancements in causal inference techniques, such as instrumental variables (IVs) and difference-in-differences (DiD), have enhanced analytical methods. However, many studies, especially those utilising large digital platform datasets, continue to encounter difficulties in achieving reliable identification within digital economy contexts [12-14]. A review of the existing literature suggests that RR's bias analytical techniques are still underutilised both as a diagnostic tool in the preliminary stages of econometric digital data analysis and as a method for mitigating bias in financial and economic research [15-17]. Additionally, there seems to be a scarcity of studies that specifically apply RR to data from the digital economy's platform datasets. Consequently, this research is essential in addressing the identified gap by assessing the use of RR as a diagnostic approach for detecting measurement errors and endogeneity in econometric models, with a particular focus on empirical data analyses within the digital economy.

Literature Review

Empirical Literature Review

The concept of using reverse regression (RR) goes back to the work of Cochrane and Hausman [3,4]. They suggested that examining how future variables relate to current ones can significantly help us understand equilibrium relationships and where predictability originates. The interest in RR has been rising exponentially because it can identify classical measurement errors, which occur when an explanatory variable is not measured accurately. RR has attracted attention in the field of econometrics, economics and applied finance due to its ability to detect classical measurement error, which arises when an explanatory variable is measured inaccurately [3]. In traditional settings, attenuation bias in ordinary least squares (OLS) estimates can be identified by comparing coefficients derived from direct and reverse regressions. A significant difference between these coefficients suggests the presence of measurement error. Additionally, RR has been employed to identify simultaneity bias in models where the direction of causality is ambiguous [18]. For example, in supply and demand models, RR helps determine whether price affects quantity or if the causal relationship operates in the opposite direction.

Hu develops identification results for regression models with mismeasured covariates and demonstrates how bounds on structural parameters can be recovered, even when regressors are observed with error, a pervasive problem in modern digital and administrative datasets [19]. The book formalises conditions under which one can recover upper and lower bounds on true regression effects by effectively "reversing" the regression and exploiting observable covariance structures between noisy proxies and latent variables. This line of work demonstrates that, under mild symmetry and independence assumptions, researchers can still obtain informative causal statements despite measurement error, weakening the reliance on classical instrumental-variable strategies. Hu's contribution is especially relevant for reverse regression diagnostics because it frames reverse modelling not just as a robustness check, but as a pathway to identify causal effects when clean regressors do not exist.

Wei and Wright (2013) contribute to the body of research focused on long-term forecasting, particularly in the context of predicting asset returns and macroeconomic indicators through reverse regression techniques [20]. These long-horizon regressions, which entail forecasting a variable such as stock returns or GDP over multiple future periods, often encounter obstacles like small-sample bias, overlapping data points, and suboptimal performance in out-of-sample testing. Wei and Wright investigate the potential benefits of inverting the regression that predicts future values based on current variables, examining whether this approach can yield improved statistical properties or new insights.

Zhang and Cheon expand the reverse regression framework to address trials characterised by nonignorable missingness, a scenario in which dropout is associated with unobserved current health status. They demonstrate that by reversing the regression, predicting treatment assignment based on observed outcomes, it becomes feasible to assess treatment effects without the necessity of fully defining the missingness mechanism, provided that dropout is conditionally independent of treatment after accounting for observed outcomes. This approach facilitates credible inference regarding treatment differences, even in instances where conventional missing-data assumptions (such as missing at random) are not upheld. The research illustrates that reverse regression can thus function as a robust tool in practical causal evaluation, particularly when traditional likelihood-based or selection-model corrections are unreliable.

Otero and Baum investigate the robustness and effectiveness of unit root tests, specifically the Dickey–Fuller (DF) and augmented Dickey–Fuller (ADF) procedures, by proposing a framework that integrates both forward and reverse regression methodologies [21]. Their objective is to enhance the statistical power and depth of these tests, thereby contributing to more reliable econometric analysis.

Dufour and Kang reexamine the theory of reverse regression (RR) within the framework of the classical linear regression (CLR) model, highlighting the distributional symmetry that exists between forward and reverse regressions under a Gaussian joint distribution [17]. Although frequently regarded as a technique for rectifying measurement errors, their research positions RR as a 'reflection' of forward regression, providing novel perspectives on its inferential characteristics and hypothesis testing methodologies. By formalising RR, they broaden its applicability beyond mere error correction, emphasising its importance in preliminary econometric analysis, particularly under assumptions of multivariate normality.

Schaefer and Visser add to the ongoing academic discussion surrounding reverse regression within the context of employment discrimination analysis, with a specific focus on wage and salary differentials. Their research represents a practical contribution to the domains of applied econometrics and forensic economics, investigating how different regression techniques—specifically forward regression, reverse regression, and orthogonal regression yield varying conclusions regarding the existence and magnitude of discrimination. The authors argue that orthogonal regression offers a more balanced approach by alleviating the arbitrary asymmetry present in both forward and reverse regression methodologies.

Greene evaluated RR methodologies within the context of wage discrimination analysis, particularly emphasising labour economics. The study by Greene presents an algebraic and statistical examination of the implications arising from the interchange of dependent and independent variables in discrimination research [22]. He illustrates that reverse regressions are not simply symmetrical counterparts to forward regressions; the estimated coefficients differ due to variations in variance, covariance structures, and group sizes. The paper includes explicit algebraic derivations that demonstrate reverse regression produces estimates of discrimination that are both statistically and economically distinct from those derived from forward regression, rather than being mere re-expressions of the same fundamental relationship. The author explicitly cautions against interpreting RR as a reliable method for measuring bias, especially in the absence of robust assumptions concerning homoscedasticity, normality, and linearity.

Ash provides a critical examination of the application and interpretive difficulties associated with reverse regression techniques within the legal profession, particularly in the context of employment discrimination cases [15]. The chapter analyzes the increasing dependence of defendants on reverse regression (RR) as a strategy to challenge discrimination claims, emphasizing the propensity for such techniques to rely on flawed or inappropriate logical premises. Ash highlights significant methodological issues, particularly the possibility that reverse regressions may contravene essential linear regression assumptions such as exogeneity and accurate model specification [15]. The research cautions that courts lacking familiarity with the statistical constraints of RR may misinterpret its findings, potentially resulting in erroneous judgments in legal contexts.

Chen explores the effectiveness of forward (direct) and reverse regressions in estimation, utilizing both ordinary least squares (OLS) and instrumental variables (IV) methodologies [13]. The study concentrates on assessing returns to scale and technological advancements within the U.S. manufacturing sector over a period of roughly fifty years. Operating within an error-in-variables (EIV) framework, where both input and output growth rates are influenced by measurement error, the research reveals that OLS estimates, whether direct or indirect, are inconsistent, with indirect OLS generally demonstrating higher precision under the assumption of normality. In summary, the study highlights the considerable bias inherent in OLS forward and reverse regressions due to measurement error and confirms that reverse IV estimation effectively mitigates this bias, ensuring consistency across both regression types.

Phillips and Shi make a significant contribution to the fields of econometrics and financial economics by presenting reverse regression as an innovative technique for identifying and timing speculative financial bubbles [23]. This method contrasts with traditional prediction approaches by regressing current prices against future prices to assess explosive behaviour, asserting that it is less susceptible to measurement errors compared to standard methods. Their reverse regression methodology enhances the real-time monitoring of bubbles, thereby improving the reliability of both detection and dating, even in the face of challenges such as endogeneity and measurement inaccuracies in financial data. The theoretical framework they propose effectively addresses the endogeneity problems characteristic of reverse regressions, reinforcing its position as a practical and robust approach for analysing financial bubbles. In conclusion, Phillips and Shi emphasize the benefits of reverse regression in mitigating bias arising from measurement errors and endogeneity, resulting in more precise detection and dating of financial bubbles [23].

The Theoretical Framework Underpinning Reverse Regressions

In outlining the theoretical underpinnings of reverse regressions, the study follows the notations in Cochrane, Hausman and Hanssen [3,4,14]. It is understood that the notations align with those found in standard econometric literature. This methodology is thought to enhance clarity when presenting models, including the correction of measurement error and reverse regression analyses, which are well-established in the field of econometrics and related applied studies. The use of Reverse regression originates from the econometric problem of measurement error and endogeneity that arise when explanatory variables are imperfectly measured or simultaneously determined with the dependent variable. Classical regression assumes that the independent variables are measured without error and are exogenous. However, in real-world digital data sets where data is often self-reported, algorithmically transformed or aggregated, the classical linear regression (CLR) assumptions rarely hold. The violation of these assumptions or conditions leads to biased and inconsistent parameter estimates in ordinary least squares (OLS) estimation, a phenomenon known as attenuation bias [3,4].

Consider the standard CLR model in equation 1:

$$Y=\alpha+\beta X+\varepsilon, \tag{1}$$

Where X is measured with error, the observed regressor then becomes

$$X^{*}=X+u \tag{2}$$

With u representing measurement error (ME) uncorrelated with Y(dependent variable) and X (an independent variable). Estimating the regression of Y on X^{*} produces a biased estimate of β given by equation 3:

$$\widehat{\beta_{OLS}} = \beta\lambda, where\, \lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} < 2 \tag{3}$$

This coefficient shrinkage reflects attenuation toward zero, resulting in underestimation of the true causal effect. This means that the OLS estimator of β is biased towards zero. In high-dimensional settings, regularised forms of ordinary least squares, including ridge regression, impose penalties on the coefficients to reduce their magnitude, thus lowering variance. While this procedure does not represent 'bias' in the conventional sense, the estimated beta coefficients are intentionally biased toward zero to enhance the optimisation of the mean squared error, MSE (Hausman, 2001).

Given the foregoing sources of attenuation bias, reverse regression (RR) addresses this issue by inverting the roles of the dependent and independent variables:

$$X_i = \phi_1\, Y_i+\psi_i \tag{4}$$

If X is measured with error, the reverse regression slope γ_1 can be compared to the inverse of β [3]. The slope γ_1 will differ in magnitude and direction depending on the presence and nature of the measurement error or endogeneity

Methodology

The dataset employed in this research was constructed to replicate the typical characteristics of digital economy data that are subject to measurement errors. Specifically, the data exemplifies a theoretical online business focused on digital sales. This study conceptualises a scenario involving the digital sales data of an online medium-sized enterprise, which is associated with advertising expenditures reported through Google or Meta dashboards. This paper investigates the effect of digital advertising expenditure on digital sales. The dataset used in this research was simulated to reflect the common traits of digital economy data, which are often impacted by measurement errors. In particular, the data illustrates a hypothetical online company centred on digital sales. The research envisions a situation involving a medium-sized enterprise's digital sales data linked to advertising spending reported via Google or Meta dashboards. The study assesses the influence of digital advertising spending on digital sales. This dataset encompasses the actual digital advertising expenditures that are not observable in practice (following the model proposed by), the reported digital advertising spending with measurement error, the digital sales figures with the measurement error added to both the actual and reported digital advertising expenditures, and the random error in the digital sales. Utilising conventional econometric simulation techniques , the true independent variable, reported digital advertising expenditure, was sampled from a normal distribution [24]. Subsequently, measurement errors were systematically incorporated to replicate the noisy reporting typical of digital datasets. The dependent variable, digital sales, was generated as a linear function of the true independent variable with the addition of random error. This methodology establishes a controlled setting to assess the efficacy of reverse regression as a diagnostic and corrective tool in the presence of realistic data noise, which is frequently encountered in digital economy data analyses [20,25].

Estimating the Attenuation Bias

Model 1: Forward/Direct OLS Regression

$$Y = \alpha_0+\beta_1 X+u \tag{5}$$

Model 2: Reverse or Backward Regression

$$X = \phi_0+\phi_1 Y+\psi \tag{6}$$

If X is measured with error, the reverse regression slope, ϕ_1 can be compared to the inverse of β (Cochrane, 2001). This means that the bias indicator can be estimated as shown in equation 7:

$$\text{Bias Indicator} = \frac{|\widehat{\beta_1}|}{|\widehat{\phi_1}|} \tag{7}$$

The ratio obtained in equation 7 is a measure of the degree as well as the direction of the attenuation bias or measurement error (ME)[4]. The ME is defined as:

$$x_i = x_i + v_i^{*}, where\, v_i^{*} \sim N(0,\sigma^2) \tag{8}$$

The simulation strategy involves executing both direct and reverse regressions on the simulated dataset, followed by an analysis of the estimated slopes and their statistical significance.

The expectations from the attenuation bias are:

- When *Bias* = 0; implies that both forward and reverse regressions agree or align with each other
- If *Bias* > 0, implies that the forward regression estimate is attenuated. It also means that the ratio differs from the reciprocal of the regression slope.

The Simulation Strategy

This section outlines the extent to which the simulated environment in this study replicates actual digital data from the digital economy and the measures implemented to ensure external validity. To enhance external validity, a variety of methodological strategies were employed. First, the measurement error (ME) was calibrated to correspond with results from empirical research, guaranteeing that the noise neither amplifies nor diminishes the real-world challenges associated with digital reporting [26]. A systematic introduction of measurement error was applied to the reported advertising expenditure (referred to as Ad Spend) to accurately depict the noisy reporting and attribution gaps typically observed in digital data from platform dashboards, such as Amazon, Google Ads and Meta [27]. Second, the parameters for mean and variance were established based on industry standards for digital small and medium-sized enterprises (DSMEs), ensuring that the findings are relevant and applicable to real-world business contexts. Third, the simulation model was executed with multiple random seeds to mitigate the possibility that the outcomes were influenced by a single data sample. Fourth, residual diagnostics were conducted to verify compliance with classical regression assumptions [28]. Collectively, these measures ensured that the simulation environment accurately reflects realistic digital data conditions while preserving analytical tractability, thereby enhancing confidence that the reverse regression (RR) results are both robust and relevant to practical digital economy scenarios.

Results and Discussions

Figure 1 illustrates the scatter plot depicting the relationship between digital sales and digital advertising expenditure. The plot clearly indicates a strong positive correlation between reported online advertising and digital sales.

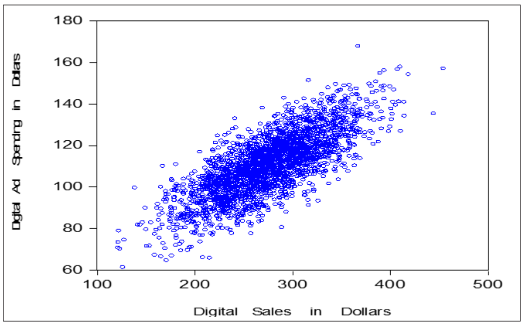

Figure 1: Scatter Plot Digital Advert Expenditure Vs Digital Sales

Expenditure on online advertising is closely associated with increased digital sales, as it enhances product visibility and consumer awareness. Advertisements enhance brand recognition and inform consumers about product features and promotions, thereby motivating them to make informed purchasing decisions. Targeted advertising campaigns effectively reach particular demographics, which enhances both cost efficiency and overall effectiveness. Interactive and personalized advertisements engage consumers on a direct level, leading to higher conversion rates. The immediate access afforded by digital platforms enables users to transition effortlessly from viewing an advertisement to finalising a purchase [29,30]. Table 1 reports descriptive statistics of the simulated data.

Table 1: Descriptive Statistics

	Mean	Max	Min	Std. Dev	Obs.
Dig_Sales	252.078	3324.16	1051.24	38.66	2433
True_Ad_Spend	97.11	109.91	51.09	11.02	2433
Reported_Ad_Spend	88.12	143.67	60.06	19.59	2433
Ran_Error_Sales	-1.111	88.17	-105.77	28.03	2433
Mean Error	-0.312	41.38	-29.67	11.53	2433

Notes: Dig_Sales denotes digital sales in millions of USD, True_Ad_Spend denotes True Advertising expenditure, Reported_Ad_Spend denotes reported online advertising spending, Ran_Error_Sales denotes the Random error added to the digital sales variable to mimic real-world noise data.

In Table 2, the justification or rationale for the distribution and the error terms in the used simulation strategy is provided.

Table 2: Justification for the Distribution and Error Terms Employed in the Simulation

Variable	Choice of Distribution Error or Error Term	Rationale for Choosing the Term	Effect on Generalization
True Digital Advertising Spending	Norma distribution	Illustrates the standard advertising expenditure patterns observed in medium-sized companies; characterised by symmetry and ease of parameterization; extensively utilized in econometric simulations (Zeng et al. 2008)	Effectively applies to typical or medium-sized companies that exhibit moderate fluctuations in advertising spending; however, it may fail to adequately capture outliers and irregular spikes in expenditure, such as those occurring during seasonal promotions.
Reported Advertising Spending	True Advert spending with systematic measurement error	Reflects real-world discrepancies that exist between actual and reported data on platforms such as Google and Meta; these discrepancies resemble attribution errors, tracking deficiencies, and reporting biases.	Enhance external validity by replicating noisy digital datasets. Generalizability depends on how closely the simulated error matches real-world reporting errors.

Digital Sales As Dependent Variable	Linear function of true sales expenditure with random error (normal, i.i.d)	Accounts for unobserved disturbances, including competitor activities and seasonal variations. This aligns with the foundational assumptions of classical regression (Wei & Wright, 2013)	Facilitate accurate estimations and diagnostic evaluations. Nonetheless, this approach may overly simplify markets characterised by heteroscedasticity or structural breaks, consequently restricting its relevance to highly volatile digital contexts.
Measurement Error in the True and Reported digital spending	Additive, normally distributed	Measurement error is systematically introduced to control the magnitude of noise, thereby facilitating the evaluation of reverse regressions under controlled conditions (Ryan and Yang, 2015)	Enhances internal validity; however, it may not completely generalize if real-world errors exhibit non-Gaussian characteristics or are correlated with actual spending.
Random Error in Sales	Normal, mean zero, homoscedastic.	Produces unbiased residuals, enabling standard econometric inference	Appropriate for relatively stable markets but may require robust adjustments when applied to data with volatility clustering or seasonality.

The study reports on the correlation relations among the variables in Table 3.

Table 3: Correlation Matrix

	DIG_SALE	T_AD_SPEND	REPORT_AD_SPEND	RAN_ERROR	ME
DIG_SALES	1				
T_AD_SPEND	0.654	1			
REPORTED_AD_SPEND	0.716	0.943	1		
RAN_ERROR	0.706	0.003	0.004	1	
ME	-0.007	-0.006	0.671	0.009	1

Notes: DIG_SALES denotes the digital sales variable, T_AD_SPEND refers to True advertising spending or expenditure, RAN_ERROR denotes random error, and ME denotes Measurement error.

The correlation matrix in Table 3 indicates a strong and positive relationship between digital sales and both total advertising expenditure (r = 0.654) and reported advertising expenditure (r = 0.716). This suggests that an increase in advertising spending is likely to enhance online sales performance. Furthermore, the reported advertising expenditure demonstrates a very strong positive correlation with the true advertising expenditure (r = 0.943), implying consistency between the actual and reported data. In contrast, random error exhibits a strong positive correlation with digital sales (r = 0.706), suggesting the presence of unobserved factors that may influence both variables, although it remains uncorrelated mainly with other variables. Measurement error (ME) exhibits a weak or negative correlation with most variables, except for reported advertising expenditure (r = 0.671), where a moderate positive relationship suggests potential inaccuracies or biases in reporting. In summary, the findings suggest that advertising expenditure effectively drives digital sales, although reported figures may be subject to measurement bias.

In Table 4, the study reports the findings obtained from both direct (forward) regressions and reverse (RR) regressions on simulated data, using Equations 5 and 6, respectively. Each variable was first differenced, as the unit root tests revealed that all variables were non-stationary at their levels. To conserve space, the unit root test results are not reported here.

The findings presented in Table 3 suggest that both actual and reported expenditure on digital advertising have a positive and statistically significant impact on digital sales in the forward regressions (Model 1). Holding other variables constant, a one-unit increase in true digital advertising expenditure results in a 5.04% increase in digital sales, whereas a one-unit rise in reported advertising expenditure leads to a 4.78% increase in digital sales. These findings validate the notion that investments in advertising significantly boost online sales performance. Likewise, in the reverse regressions (Model 2), both true and reported digital advertising expenditures continue to exhibit positive and significant correlations with digital sales, with coefficients of 3.06% and 2.72%, respectively. This consistency across different models highlights the strength of the advertising-sales relationship, suggesting that both accurately measured and reported expenditures stimulate consumer engagement and purchasing behaviour.

The positive and significant impacts can be linked to the capacity of online advertisements to enhance product visibility, foster brand awareness, and draw potential customers to e-commerce platforms. As noted by Grewal et al., digital advertisements inform consumers about product features and promotions, while interactive and personalised formats, such as clickable ads and real-time feedback, improve engagement, trust, and ultimately, purchase decisions [31,32]. In summary, the results underscore the crucial role of digital advertising in shaping consumer behaviour and promoting sustained sales growth.

Table 4: Results from Forward and Reverse OLS Regressions

	Coeff	Std.Error	t-Stat	P_value	Mean Dependent Var
Model 1: Direct OLS Regressions (Method: Least Squares)					
Model 1A: Regressant- DIG_SALES with Ran_Error					
Regressor- T_AD_ SPENDING	5.042**	0..411	21.872	0.001	302.145
Model 1B: Regressant- DIG_SALES with Ran_Error					
Regressor-REPORT_AD_SPENDING	4.777**	0.219	-17.051	0.002	266.095
Model 2: Reverse Regressions					
Model 2A: Regressant- DIG_SALES with Ran_Error					
Regressor- T_AD_SPENDING with Ran_Error	3.061**	0.005	21.804	0.003	110.086
Model 2B: Regressant-DIG_SALES with Ran_Error					
Regressor-REPORTED_AD_SPENDING	2.717**	0.007	13.432	0.004	99.889

Source: Author's Elaboration on Data. ** Denotes significant at 5% level.

Notes: DIG_SALES denotes the digital sales variable, T_AD_SPEND refers to True advertising spending or expenditure, and RAN_ERROR denotes random error.

Calculating the Bias Indicator

In this section, the bias indicator is estimated by employing equation 7:

Bias Indicator = $\frac{|\widehat{\beta}_1|}{|\widehat{\phi}_1|}$

The coefficients in Model 1A and 2A are compared, whereas the coefficient in Model 1B is compared to those in Model 2 B.

Model 1A and 2A compared (Direct LS Regressions): Bias Indicator = $\frac{|\widehat{\beta}_1|}{|\widehat{\phi}_1|}$ **= 5.04/3.06 = 1.65**

Model 1B and 2 B compared (Direct LS Regressions): Bias Indicator = $\frac{|\widehat{\beta}_1|}{|\widehat{\phi}_1|}$ =4.78/2.72 = 1.76

The bias ratio, or bias indicator, is found to be positive in both Model 1 and Model 2. This indicates that the forward (direct) estimates derived from forward regressions are skewed towards zero (attenuated). In practical terms, this underestimation may result in businesses underinvesting in advertising, erroneously concluding that their digital campaigns are less effective than they really are. For policymakers, attenuation bias could distort evaluations of the role of digital marketing in economic growth, potentially resulting in insufficient support for digital transformation initiatives or misaligned incentives for SMEs within the digital economy. For instance, if sales elasticities are underestimated, tax policies or innovation grants associated with digital adoption may not be appropriately adjusted to reflect their genuine economic impact [33,34]. Furthermore, inaccurate estimates also compromise the formulation of evidence-based marketing strategies, as companies might divert budgets away from high-performing channels. Therefore, recognising and addressing attenuation bias through methods such as reverse regression has substantial implications for optimising advertising expenditures, enhancing policy targeting, and ensuring that data from the digital economy accurately informs decision-making [35].

Table 5 reports the results from sensitivity analyses in line with the simulation strategy described in section 3.3.

Table 5: Key Results from the Sensitivity Analyses

Test condition	Findings	Comment
Measurement-Error Variance	Reverse regression (RR) shows consistently lower root mean squared error (RMSE) than the forward/direct OLS. The gap widens as measurement error (ME) increases (See Figure 1-7 in Appendix 3)	In econometrics, a general guideline is that a reduced RMSE signifies enhanced model accuracy, suggesting that predictions align more closely with the actual values (or that sample parameters are more representative of population parameters). In contrast, an elevated RMSE indicates reduced model accuracy, as predictions diverge further from the actual values.

Error Distributions	RR retains RMSE advantage under non-Gaussian errors (t, lognormal, Laplace). See Figures 2 and 3 in Appendix 3.	These results highlight the strength of Ridge Regression (RR), as it continues to yield lower Root Mean Square Error (RMSE) than Ordinary Least Squares (OLS) even when the error terms do not conform to a normal Gaussian distribution. This is significant since actual data seldom adheres to ideal normality. It suggests that RR can provide more dependable estimates in real-world contexts, such as data from digital platforms, where noise frequently diverges from Gaussian assumptions.
Endogeneity Stress	Bias increases for both Forward and Reverse Regressions as ρ(X_(true,) u) rises. Refer to Figure 4 in Appendix 3	This underscores that endogeneity affects both Reverseand Forward OLS. It means that as the correlationbetween the true independent variables (denoted X_true) and the error term,μ. Both methods produce more biased estimates. RR alone cannot address endogeneity problems. To address this empirical problem, RR should be paired with techniques like instrumental variables (IV) or difference-in-difference (DID). These methods help isolate the causal effects of X_true by breaking its correlation with the error term, leading to more reliable and consistent estimates.
Heteroscedastic ME	RSME rises for both methods, but RR maintains relative efficiency. See Figure 5 in Appendix 3	Although both estimators deteriorate as heteroscedasticity increases, the estimates produced by RR tend to be closer to the actual parameters on average. This indicates that RR is more efficient and better able to withstand noisy, uneven data.
Platform Drift	Mean elasticities shift under β change. This highlights the importance of drift diagnostics. See Figures 6 and 7 in Appendix 3	This result signifies that when the essential relationship between the variables represented by β changes, whether due to platform updates, market fluctuations, or alterations in user behaviour, the average elasticities also shift. This implies that models relying on historical data may become misaligned over time if these changes are not detected and corrected.

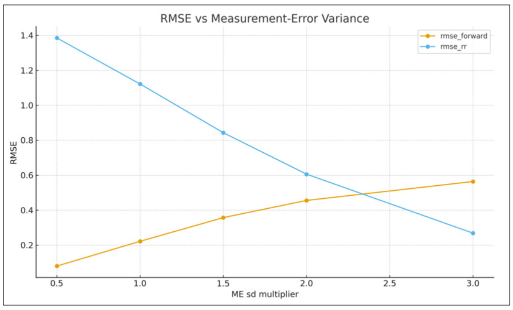

Figure 1: RMSE vs Measurement-Error Variance

Notes: Simulated data calibrated to the provided descriptive statistics (n=2433 baseline). R=500 replications per condition. RR denotes a symmetric reverse-regression estimator (GMFR), Forward denotes OLS of sales on reported ad spend. The measurement-error standard deviation is multiplied by {0.5, 1.0, 1.5, 2.0, 3.0}.

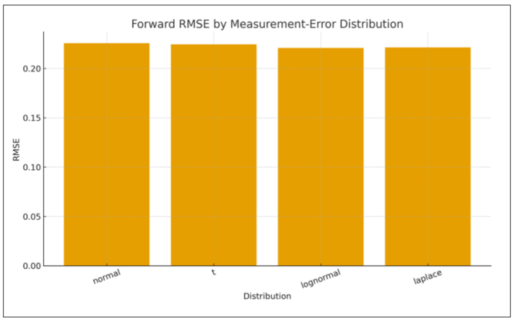

Figure 2: Forward RMSE by Measurement-Error Distribution

Notes: R=400 per condition. Distributions: Normal, Student t(df=5; heavy tails), Lognormal (skewed), Laplace (double-exponential).

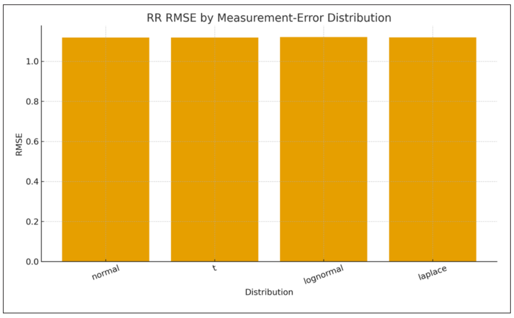

Figure 3: Reverse-Regression (GMFR) RMSE by Measurement-Error Distribution

Notes: Same design as Figure 2. RR shows relative RMSE advantages under non-Gaussian errors in these simulations.

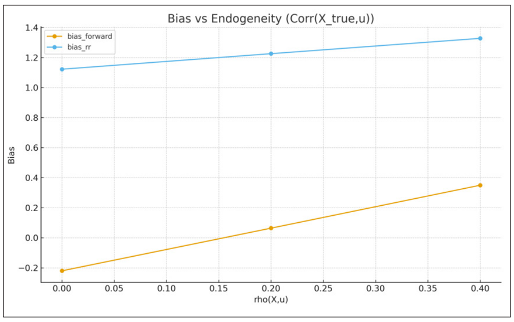

Figure 4: Bias vs Endogeneity (Corr(X_true, u))

Notes: R=500 per condition. Endogeneity introduced via correlation $\rho \in \{0, 0.2, 0.4\}$ between true ad spend and sales shocks. Both estimators' bias increases with ρ, motivating RR+IV/DID hybrids.

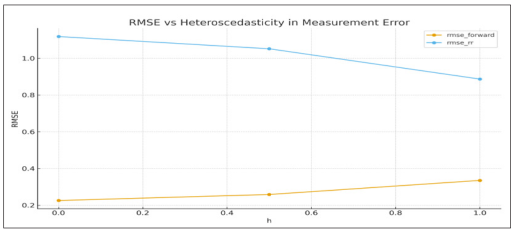

Figure 5: RMSE vs Heteroscedasticity in Measurement Error

Notes: R=400 per condition. Heteroscedasticity parameter $h \in \{0, 0.5, 1.0\}$ scales error variance with X_true. RR maintains relative efficiency in these tests.

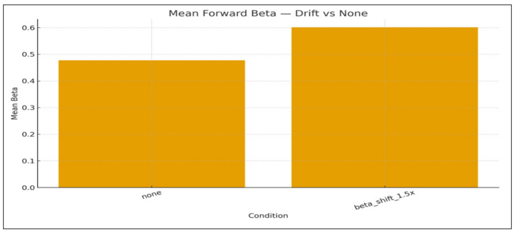

Figure 6: Mean Forward Elasticity — Platform Drift (β shift)

Notes: R=400 per condition. Drift scenario: post-break β = 1.5× baseline vs no drift. Highlights estimator sensitivity to structural breaks.

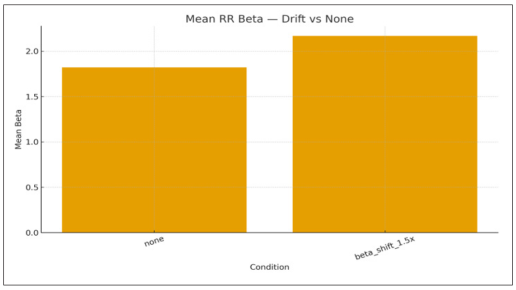

Figure 7: Mean RR Elasticity — Platform Drift (β shift)

Notes: Same design as Figure 6, for the RR estimator

General Notes: Baseline n=2433; replications per block as indicated. RR = geometric mean functional relationship (reduced major axis). Forward = OLS of Y on X_reported. Results are simulation-based; coverage reported for forward OLS only.

Comparative Literature Review: Application of Reverse Regressions

Reverse regression (RR) has evolved from a specialised econometric curiosity to a multifaceted diagnostic and estimation instrument applied across various disciplines. Foundational works by Griliches and Hausman position RR mainly as a remedy for attenuation bias that arises from measurement errors in explanatory variables. In this capacity, RR is frequently combined with instrumental variables (IV) techniques to obtain consistent parameter estimates in the presence of noisy inputs. Research conducted by Cready, Hurtt, and Seida demonstrates that RR can enhance inference in earnings-return models where accounting data is affected by transitory noise, while Zeng et al. illustrate that RR provides less biased calibration under heteroscedastic conditions, rendering it appropriate for scenarios where error variance correlates with the size of observations, a characteristic often found in digital platform data. In addition to addressing measurement errors, RR has been employed to investigate simultaneity and causal direction. Angrist and Pischke highlight its diagnostic significance in supply-demand frameworks, where price and quantity may mutually affect one another. Green and Schaefer and Visser argue that misinterpreting reverse and forward regressions as simple algebraic inverses can result in erroneous conclusions, as variations in variances and covariances produce economically distinct coefficients. Their warnings resonate with Goldberger's caution that RR should be regarded as a diagnostic tool rather than a structural estimator unless rigorous OLS assumptions are satisfied.

In the realms of time series analysis and forecasting, the use of RR offers significant advantages. Wei and Wright demonstrate that the application of flipping regressions can enhance the stability of inferences in long-horizontal predictive models, particularly when the predictors exhibit persistence or overlap. In a similar vein, Phillips and Shi utilise RR to identify speculative bubbles within financial markets, claiming that regressing current prices against future prices reduces measurement errors and improves the accuracy of real-time bubble dating methodologies. Cartwright and Riabko present a crucial caveat for practitioners, arguing that while temporal aggregation, such as transitioning from daily to monthly data, can enhance the explanatory capacity of regression models, it may also lead to inflated standard errors, potentially resulting in weakened inference. This caveat is particularly pertinent for platform data that is aggregated for reasons of privacy or storage efficiency. Furthermore, recent methodological advancements have broadened the computational and diagnostic applications of RR. Dereziński and Warmuth propose Reverse Iterative Volume Sampling (RIVS), which facilitates unbiased subset selection for large-scale regressions while maintaining least squares optimality, a feature that is especially beneficial for high-dimensional datasets in the digital economy. Otero and Baum merge forward and reverse ADF tests to enhance unit-root diagnostics, thereby providing a more robust toolkit for time series applications, particularly in small sample contexts.

In this comparative analysis, three key themes are identified. Firstly, RR serves as a highly effective diagnostic tool for detecting potential measurement errors or recognising simultaneity issues that could jeopardise the results of forward regression. Secondly, the robustness of RR is enhanced when it is utilised in conjunction with complementary methods such as IVs or GMM, which address endogeneity and restore consistency. Thirdly, the applicability of RR is context dependent. Although RR can enhance inference in environments characterised by noise, heteroscedasticity, or persistence, it is advisable to present it alongside forward regression results rather than as a substitute. In the realm of digital-economy research, these themes collectively advocate for a triangulated methodology in addressing forward versus reverse gaps, serving as a tool for attenuation or bias screening. It is essential to integrate RR-IV or GMM to confirm parameter stability and explicitly assess sensitivity to temporal aggregation and heteroscedasticity. This positions RR not as a competitor to forward regression estimation but as a complementary methodological framework that equips researchers, policymakers, and practitioners with deeper insights prior to making structural or causal interpretations.

The distinctive contributions of this study can be enumerated as follows. The research provides a thorough synthesis of the development and applications of reverse regressions (RR), establishing them as essential instruments in contemporary econometric practice. Furthermore, the contextualization of RR emphasises that it should function as a supplementary diagnostic tool to forward regressions rather than a direct substitute. A significant contribution of this study is its contextualization of RR within the digital-economy research landscape, where data frequently exhibit noise, heteroscedasticity, and temporal aggregation. In such contexts, RR provides a robust mechanism for identifying potential distortions that could compromise the reliability of OLS estimates, thereby offering researchers and policymakers valuable insights.

The paper advocates the utilisation of RR as an initial diagnostic tool to identify possible measurement errors and simultaneity problems, subsequently enhancing it with forward regressions and robustness assessments. Specifically, it is recommended that practitioners compute and analyse forward–reverse coefficient discrepancies to uncover attenuation bias, implement RR in scenarios where predictor persistence could skew forward inferences (for instance, in Ad spending and click-through data), and utilise RR in conjunction with IVs or GMM to mitigate endogeneity challenges. The research provides explicit implementation recommendations to guarantee the effective application of RR. Initially, researchers are cautioned against over-interpreting RR coefficients as structural estimates unless the assumptions of linearity, homoscedasticity, and normality are adequately met (Goldberger, 1984; Greene, 1984). Furthermore, RR should be performed on disaggregated data whenever possible, as temporal aggregation can lead to inflated standard errors and compromised inference. These insights provide a framework for researchers and policymakers, emphasising the importance of reporting RR alongside forward regression outcomes to enhance transparency, strengthen the robustness of inferences, and inform decision-making in applied economic, financial, and digital platform environments.

Notwithstanding its contributions, this study is not without limitations. Firstly, it is based on a controlled simulation environment, which may oversimplify the intricacies of real-world digital data characterised by platform-specific noise, absent values, and unobserved heterogeneity. Secondly, the analysis is confined to a static, single-firm context, which may neglect competitive dynamics and feedback mechanisms that affect advertising choices in actual digital economy markets. Thirdly, the presumption of access to "true" advertising expenditure, while beneficial for isolating attenuation bias, may not accurately represent practical scenarios where such data is seldom available. Moreover, the study does not incorporate RR with alternative causal inference methodologies such as IVs or DID, which restricts its generalisation. In light of the foregoing findings, the study recommends the following frameworks and future studies.

Complex Error Structures

Future research should aim to simulate non-Gaussian and heteroscedastic error structures to represent real-world data from digital platforms accurately. By integrating platform-specific biases and correlated errors, researchers can assess the robustness of reverse regression in practical scenarios and determine the conditions under which RR surpasses forward regression in efficiency.

Dynamic and Competitive Market Contexts

This investigation focused on a single firm within a static environment, which constrains its breadth. Future inquiries ought to employ panel or longitudinal datasets to observe the evolution of bias correction over time. Macroeconomic and Policy Implications: Future studies should investigate how bias-adjusted advertising elasticities impact GDP, productivity, and contributions from various sectors. The refined metrics can subsequently inform the development of more effectively targeted digital subsidies and tax incentives.

Robustness and External Validity

Future research should utilise data that is replicated across diverse industries, platforms, and geographical areas to affirm the generalizability of reverse regression findings.

Development of Practitioner-Oriented Tools

To achieve a broader impact, future studies should contemplate the translation of reverse regression into practical decision-support instruments, such as R or Python packages, dashboards, or application programming interfaces (APIs) that could automate bias estimation and assist firms in evaluating advertising return on investment (ROI). Such tools would enhance the accessibility of reverse regression for small and medium-sized enterprises (SMEs), effectively bridging the gap between theory and practice.

Appendices

Appendix 1: Simulation Design and Modeling Strategy

To ensure that the results are reproducible and reflect real-world conditions, all baselines were adjusted to correspond with the empirical traits of digital data. In particular, the research synchronised the means and standard deviations of the variable (Y representing the digital sales with the random error added to the digital sales), the true predictor (X_true representing true digital advertising spending), the reported predictor (X_rep representing reported digital advertising spending with measurement error), the structural disturbance term (u), and the random error component (ε). The sample was set at n= 2433 to attain a balance between statistical power and computational practicality. The simulation scripts were crafted to methodically investigate a variety of data-generating scenarios, facilitating a comprehensive stress test of the estimators. The scenarios differed across multiple essential dimensions:

- Measurement-Error Variance and Distributions: Noise in X_rep was manipulated across Gaussian, heavy-tailed (t), skewed (lognormal), and Laplace distributions to test robustness.
- Heteroscedasticity: Error variance scaled with X_true, mimicking real-world data patterns.

- Endogeneity Stress Tests: Correlation between X_true and u (ρ(X_true,u)) was varied to observe estimator behavior under exogeneity violations.
- Signal-to-Noise Ratio (SNR): Different SNR levels were tested to measure performance under weak signals.
- Sample Size Variations: Smaller and larger n were simulated to evaluate scalability.
- Platform Drift: Both β-shifts and ME variance shifts were incorporated to emulate changes in underlying data-generating processes.

Two estimation strategies were implemented and compared
- Forward OLS: Standard regression of Y on X_rep.
- Reverse Regression (RR) inspired Symmetric Estimator: Geometric Mean Functional Relationship (GMFR) or reduced major axis regression as a reverse regression proxy.

Performance Metrics Included:
- Mean Slope: Average estimated effect across simulations.
- Bias: Deviation of mean slope from true β.
- RMSE: Root Mean Squared Error.
- 95% CI Coverage: Reliability of inference for forward OLS.
- RR Efficiency Gain: Proportional MSE reduction relative to forward OLS.

Together, these design elements provide a comprehensive assessment of estimator performance, highlighting accuracy, stability, and robustness under varied conditions.

Appendix 2: Results of the Sensitivity Analyses
- Measurement-Error Variance: Reverse Regression (RR) consistently shows lower RMSE than forward OLS. The gap between RR and OLS widens as measurement-error (ME) variance increases (see Figure 1).
- Error Distributions: Under non-Gaussian errors (t, lognormal, Laplace), RR generally maintains its RMSE advantage over forward OLS (See Figures 2–3).
- Endogeneity Stress: When the correlation between true X and the error term increases, both methods become more biased.

This highlights the need to combine RR with IV or Difference-in-Differences (DID) methods in endogenous settings (Figure. 4).
- Heteroscedastic Measurement Error: RMSE increases for both RR and OLS as heteroscedasticity (h) rises. However, RR still preserves relative efficiency under these conditions (Figure. 5).
- Platform Drift: Mean elasticities shift when β changes, showing that both estimators are sensitive to drift. Figures 6–7 suggest that monitoring for platform drift is necessary.

Appendix 3: Sensitivity Analysis of Reverse & Forward Ols Regressions
Calibrated to the provided statistics (DIG_SAL, AD_SPEND, REPORTED_AD_SPEN, RANDOM_ERROR_SALE, MEASUREMENT_ERROR); β_true = 0.7; α chosen to match sales mean; large-sample normal CIs for OLS.

Conclusion
This research investigates reverse regression (RR) as a diagnostic instrument within the digital economy, utilising digitally simulated data from a hypothetical medium-sized online firm. The study was conducted in the context of significant data generation by digital platforms, which can be utilised by researchers and policymakers to gain insights into market dynamics, consumer preferences, and employment trends. Nonetheless, the data obtained from these platforms often suffer from measurement errors due to self-reported information, algorithmic adjustments, and inadequate validation procedures. This research presents RR as a diagnostic tool to address these econometric issues within the digital economy framework. The study aimed to assess RR's effectiveness as a diagnostic tool for identifying measurement errors and endogeneity in datasets pertaining to the digital economy. It evaluated RR's capacity to reveal bias when ordinary least squares (OLS) assumptions are not met, thereby assisting researchers and policymakers in avoiding erroneous conclusions. The findings illustrate how RR bolsters the robustness of analyses that utilise platform-generated data and enhances the reliability of parameter estimates. It underscores the importance of integrating regression analysis with instrumental variables (IVs) and the generalised method of moments (GMM) to restore consistency and strengthen causal interpretations. Furthermore, the research proposes a policy-oriented framework designed to ensure more precise calibration of tax incentives, innovation grants, and digital economy interventions. The regression results indicate that, holding other variables constant, a one-unit increase in true digital advertising expenditure results in a 5.04% increase in digital sales, whereas a one-unit rise in reported advertising expenditure leads to a 4.78% increase in digital sales. These findings validate the notion that investments in advertising significantly boost online sales performance. Likewise, in the reverse regressions (Model 2), both true and reported digital advertising expenditures continue to exhibit positive and significant correlations with digital sales, with coefficients of 3.06% and 2.72%, respectively. This consistency across different models highlights the strength of the advertising-sales relationship, suggesting that both accurately measured and reported expenditures stimulate consumer engagement and purchasing behaviour [36-40].

Funding Statement
This study received no external funding.

Acknowledgments
The Author thanks the anonymous reviewers. Their insights significantly helped to improve the quality of this paper.

Declaration of Interest
The author has no financial or non-financial conflict of interest to declare.

References
1. Goldfarb A, Tucker C. Digital economics. Journal of Economic Literature. 2019;57:3–43.
2. UNCTAD. Digital economy report 2019: Value creation and capture – Implications for developing countries. United Nations. 2019.
3. Hausman JA. Mismeasured variables in econometric analysis: Problems from the right and problems from the left. Journal of Economic Perspectives. 2001. 15: 57-67.
4. Cochrane JH. Asset pricing Princeton University Press. 2001.

5. WM, Hurtt DN, Seida JA. Applying reverse regression techniques in earnings–return analyses. Journal of Accounting and Economics. 2000. 30: 25-45.
6. Einav L, Levin J. The data revolution and economic analysis. Innovation Policy and the Economy. 2014. 14: 1-24.
7. Luca M. Reviews, reputation, and revenue: The case of Yelp.com. Harvard Business School NOM Unit Working Paper No. 2016.
8. Nosko C, Tadelis S. The limits of reputation in platform markets: An empirical analysis and field experiment. NBER Working Paper No. 20830. 2015.
9. Bound J, Brown C, Mathiowetz N. Measurement error in survey data. In: Handbook of Econometrics. 2001. 5: 3705-3843.
10. Athey S, Imbens GW. The state of applied econometrics: Causality and policy evaluation. Journal of Economic Perspectives. 2017. 31: 3-32.
11. Ghose A, Yang S. An empirical analysis of search engine advertising: Sponsored search in electronic markets. Management Science. 2009. 55: 1605-1622.
12. Greene WH. Econometric analysis. 6th ed. Springer. 2010.
13. Chen X. Increasing returns to scale in U.S. manufacturing industries: Evidence from direct and reverse regression. Working paper. Bureau d'Économie Théorique et Appliquée, University of Strasbourg; 2011.
14. Hansen LP. Uncertainty outside and inside economic models. Journal of Political Economy. 2015. 122: 945-987.
15. Ash AS. The perverse logic of reverse regression. In: Statistical Methods in Discrimination Litigation. Taylor & Francis; 2014.
16. Dereziński M, Warmuth MK. Reverse iterative volume sampling for linear regression. Journal of Machine Learning Research. 2018. 19: 1-41.
17. Dufour JM, Kang B. Reverse regressions, symmetry and test distributions in linear models. Journal of Quantitative Economics. 2022. 20: 1-25.
18. Angrist JD, Pischke JS. Mostly harmless econometrics: An empiricist's companion. Princeton University Press. 2009.
19. Hu Y. The econometrics of unobservables. Working manuscript. Department of Economics, Johns Hopkins University. 2025.
20. Wei WWS, Wright TS. Time series analysis: Univariate and multivariate methods. 2nd ed. Pearson; 2013.
21. Otero J, Baum CF. Unit-root tests based on forward and reverse Dickey–Fuller regressions. The Stata Journal. 2018. 18: 145-164.
22. Greene WH. Reverse regression: The algebra of discrimination. Journal of Business & Economic Statistics. 1984. 2: 1-8.
23. Phillips PCB, Shi SP. Financial bubble implosion and reverse regression. Econometric Theory. 2018. 34: 705-753.
24. Zeng D, Lin DY, Tu W. Measurement error models: A selective review. Statistica Sinica. 2008. 18: 905-925.
25. Ryan T, Yang H. Measurement error models. CRC Press. 2015.
26. Cartwright PA, Riabko N. Temporal aggregation and the estimation of reverse regressions for commodities market models. In: Handbook of Investment Analysis, Portfolio Management, and Financial Derivatives. 2024.
27. Chaffey D, Ellis-Chadwick F. Digital marketing. 7th ed. Pearson Education. 2019.
28. Chaffey D, Ellis-Chadwick F. Digital marketing. 8th ed. Pearson Education. 2022.
29. Chen Y, Pavlov D, Canny J. Large-scale behavioral targeting. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 .209-218.
30. Fuller WA. Measurement error models. Wiley. 1987.
31. Grewal D, Hulland J, Kopalle PK, Karahanna E. The future of technology and marketing: A multidisciplinary perspective. Journal of the Academy of Marketing Science. 2020. 48: 1-8.
32. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Data mining, inference, and prediction. 2nd ed. Springer. 2009.
33. Kydland FE, Prescott EC. Time to build and aggregate fluctuations. Econometrica. 1982. 50:1345-1370.
34. Lambrecht A, Tucker C. When does retargeting work? Information specificity in online advertising. Journal of Marketing Research. 2013. 50: 561-576.
35. Li J, O'Donoghue C. A survey of dynamic microsimulation models: Uses, model structure and methodology. International Journal of Microsimulation. 2013. 6: 3-55.
36. Meta. About Meta Ads Manager. Meta Business Help Center. 2024.
37. Neigel P, Selby DA, Arai S, Kise K. Exploring the alignment of perceived and measured sleep quality with working memory using consumer wearables. Preprint. ResearchGate. 2025.
38. OECD. Measuring digital transformation: A roadmap for the future. OECD Publishing. 2020.
39. Stephen AT. The role of digital and social media marketing in consumer behavior. Current Opinion in Psychology. 2016. 10: 17-21.
40. Wooldridge JM. Econometric analysis of cross-section and panel data. 2nd ed. MIT Press; 2010.
41. Zeng QC, Zhang E, Tellinghuisen J. Univariate calibration by reversed regression of heteroscedastic data: A case study. The Analyst. 2008. 133: 1340-1345.