How Do Sales Efforts Pay Off?
Dynamic Panel Data Analysis in the Nerlove-Arrow Framework

Doug J. Chung, Harvard University

Byungyeon Kim, Harvard University

Byoung G. Park, University at Albany

August 2018

Abstract

This paper evaluates the short- and long-term value of sales representatives' detailing visits to different types of physicians. By understanding the dynamic effect of sales calls across heterogeneous physicians, we provide guidance on the design of optimal call patterns for route sales. The findings reveal that the long-term persistence effect of detailing is more pronounced for specialist physicians, whereas the contemporaneous marginal effect is higher for generalists. The paper also provides a key methodological insight to the marketing and economics literature. In the Nerlove-Arrow framework, moment conditions that are typically used in conventional dynamic panel data methods become vulnerable to serial correlation in the error structure. We discuss the associated biases and present a robust set of moment conditions for both lagged dependent and predetermined explanatory variables. Furthermore, we show that conventional tests to detect serial correlation have weak power, resulting in the misuse of moment conditions that leads to incorrect inference. Theoretical illustrations and Monte Carlo simulations are provided for validation.

Key words: Nerlove-Arrow framework, stock-of-goodwill, dynamic panel data, serial correlation, instrumental variables, sales effectiveness, detailing, pharmaceutical industry.

1. Introduction

The pharmaceutical industry plays a significant role in the world economy. According to QuintilesIMS (formerly IMS Health), the global market for prescription drugs is expected to grow from \$1.1 trillion in 2016 to \$1.5 trillion by 2021. Despite the large size of this market, however, marketing to customers (i.e., physicians) is typically restricted to personal selling in the form of detailing by a pharmaceutical company's sales force.[1] Even in the United States, a nation that allows direct-to-consumer pharmaceutical advertising, personal selling remains the dominant marketing channel. Some 90,000 sales representatives (1 for every 6.3 doctors) market pharmaceutical products to 567,000 U.S. physicians (Wall and Brown, 2007).

Studies on the effectiveness of personal selling to generate physician prescriptions have produced strikingly mixed findings in the literature with reported sales elasticity measures ranging from −14.8% (Parsons and Vanden Abeele, 1981) to 41% (Gönül et al., 2001). This inconsistency is possibly due to both limited data on physicians' prescribing behavior and the bias arising from naive treatment of data, specifically panel data. Thus, we seek to gain insights on deriving an unbiased measure of both the short- and long-term value of a firm's detailing efforts through robust econometric analyses.

Obtaining a precise and unbiased value of detailing efforts turns out to be rather challenging. Physicians' prescribing behavior is highly habitual with a significant amount of unobserved heterogeneity. Physicians are known to exhibit a high level of inertia (Janakiraman et al., 2008), so an individual physician's past number of prescriptions is likely to persist and thus affect current sales. To accommodate this dynamic process, studies in economics and marketing have frequently adopted the Nerlove-Arrow (1962) framework, which conceptualizes sales as a function of a stock of goodwill that increases in response to a firm's current marketing activities but decays over time. To empirically model this framework, studies often have used the geometric lag model (Koyck, 1954; Balestra and Nerlove, 1966)—a form of the general dynamic panel data model

[1] As of 2016, direct-to-consumer advertising was allowed only in Brazil, New Zealand, and the United States, with varying restrictions on content.

specification—that substitutes the infinite geometric sum of marketing efforts with a lagged dependent variable.[2,3]

Identifying the causal effect of detailing becomes challenging when confronted with unobserved heterogeneity. Because pharmaceutical companies are likely to allocate more resources (e.g., shorter call cycles) to physicians with higher sales volume or growth potential, it is necessary to control for the correlation between sales efforts and potential. In addition, an endogeneity problem arises because, by construction, the lagged dependent variable is correlated with lagged error terms through unobserved heterogeneity.

Dynamic panel data methods proposed by Anderson and Hsiao (1981, 1982) and further developed by Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998) provide a practical approach to tackling the endogeneity issue while simultaneously accounting for unobserved heterogeneity. The key advantage of these methods is that they allow us to control for potential biases without relying on strictly exogenous instrumental variables, which in many empirical settings are impossible to obtain. Because of this practicality, dynamic panel data methods have been used in numerous contexts in economics and marketing, including advertising (Clark et al., 2009; Song et al., 2015; McAlister et al., 2016), customer-relationship management (Van Triest et al., 2009; Tuli et al., 2010; Rego et al., 2013), product innovation (Narasimhan et al., 2006; Fang et al., 2016), habit formation (Shah et al., 2014), entertainment marketing (Narayan and Kadiyali, 2016; Mathys et al., 2016; Chung, 2017), social media (Archak et al., 2011), marketing-finance interface (Germann et al., 2015; Feng et al., 2015), market entry (Mukherji et al., 2011), crowd funding (Burtch et al., 2013), political economics (Acemoglu et al., 2008), and growth economics (Durlauf et al., 2005).

The underlying micro-foundation of these studies largely falls into two categories: (i) those that have used the lagged dependent variable to simply control for autocorrelation with no direct

[2] The geometric lag model is sometimes referred to as the Koyck model.

[3] The framework that the geometric lag model accommodates is not limited to goodwill—e.g., stock of appliances (Balestra and Nerlove, 1966), partial adjustment (Hatanaka, 1974; Nerlove, 1958), and adaptive expectation (Cagan, 1956; Friedman, 1957). This study focuses on goodwill as it is commonly used in the literature; however, the method presented in this paper is robust and can be applied to any framework that utilizes the geometric lag model.

interpretation of the micro-foundation; and (ii) those that have explicitly or implicitly used the Nerlove-Arrow framework as the underlying micro-foundation behind the dynamics. However, under the Nerlove-Arrow framework, the use of dynamic panel data methods is afflicted by a troubling issue: the geometric lag model encompasses serially correlated errors by construction, yet the validity of conventional dynamic panel data methods relies on the assumption that the error structure does not exhibit serial correlation. If serial correlation is present and undetected, the moment conditions derived under these methods become invalid, resulting in unreliable inference. Furthermore, the predetermined nature of explanatory variables—firms observing past performance shocks to determine the current-period level of actions—poses an additional endogeneity issue.

To obtain unbiased estimates of the causal effect of detailing, we present an adequate set of moment conditions that are robust to serial correlation, in a similar vein to Hujer et al. (2005) and Lee et al. (2017). In addition, we present the means to mitigate the endogeneity concern with regard to predetermined variables, an issue that has rarely been addressed in the literature.

More importantly, we show that commonly used test statistics to detect serial correlation become biased when invalid moment conditions are used. To test for the validity of the moment conditions, and thus the model specification, past studies have relied on the Arellano-Bond test for serial correlation (Arellano and Bond, 1991). When the AR(2) test statistic is not rejected, presumably indicating an absence of serial correlation in the error structure, researchers have proceeded with the estimation without further concern. However, the Arellano-Bond specification test is prone to weak power in detecting serial correlation. The test statistic can fail to detect serial correlation and wrongly justify the use of invalid moment conditions, resulting in biased estimates and thus incorrect inference. We provide formal proof on the technical shortcomings of the Arellano-Bond specification test and specify conditions where the test can fail to reject the misspecified model. Our findings suggest that researchers be cautious about the use of conventional instruments in dynamic panel data settings, if any degree of serial correlation is suspected in the underlying model—even when the AR(2) test statistic is not rejected.

To validate our claims, we conduct Monte Carlo simulations to show that conventional methods yield biased estimates under serially correlated errors. The simulation results also reveal the weak power of the Arellano-Bond specification test using conventional moment conditions when serial correlation is present.

For the empirical application, we collaborate with a multinational pharmaceutical company and apply our method to a comprehensive panel dataset that includes detailed individual physicians' prescribing histories and the detailing efforts of the firm's sales representatives. We postulate that when serial correlation is present, conventional methods yield biased and counterintuitive estimates implying that detailing has negative effectiveness. By correcting the misuse of invalid moment conditions, the analysis reveals that detailing efforts, on aggregate, have a significant impact on physicians' prescription rates.

Subsequently, we allow for heterogeneity in the slope parameters to account for differences in the effectiveness of detailing across different medical practice areas. The results show that, in general, specialist physicians (e.g., cardiologists, diabetologists, and endocrinologists) exhibit greater persistence in prescribing patterns, whereas generalists (e.g., consulting physicians, general practitioners, and general surgeons) are more responsive to short-term detailing efforts but exhibit less persistence.

Our simple yet methodologically robust model can help firms obtain an unbiased measure of detailing efforts, which, in turn, will help firms design optimal call patterns and sales targets to increase the overall effectiveness of their sales force.

The remainder of the paper is organized as follows. Section 2 examines the Nerlove-Arrow framework and how it translates into a dynamic panel data setting; and presents our methodology that builds on conventional dynamic panel data methods. Section 3 addresses the conditions in which the test for serial correlation is prone to weak power. Section 4 presents simulation studies to verify our claims. Section 5 describes institutional details, the data, and the empirical model. Section 6 discusses the results, and Section 7 concludes.

2. Methodology

We first describe the Nerlove-Arrow framework and how it relates to a general dynamic panel data model. Next, we delineate the conventional dynamic panel data estimation methods and discuss potential bias in the presence of serial correlation. Then, we present our proposed methodology.

2.1 The Nerlove-Arrow Framework

The stock-of-goodwill framework of Nerlove and Arrow (1962; hereafter N-A) has played a pivotal role in examining the long-term effects of various marketing mixes and management practices. The key construct embodied in this framework is the formation of an unobserved stock of goodwill created by a firm's current and past actions (e.g., advertising or CRM expenditure), which affects the current period outcomes such that

$$y_{it} = G_{it} + \tilde{\gamma}' z_i + \tilde{\alpha}_i + \nu_{it} , \qquad (1)$$

where y_{it} denotes the outcome of interest (e.g., sales) and G_{it} represents the unobserved stock of goodwill for cross-sectional unit i (typically a firm or a person) at time t. The individual effect $\tilde{\alpha}_i$ represents unobserved heterogeneity that persists over time and z_i ($k_z \times 1$ dimension) denotes the vector of observed time-invariant characteristics, with marginal effects $\tilde{\gamma}$. The unobserved individual- and time-specific idiosyncratic shock ν_{it} is assumed to be serially uncorrelated over time.

The stock of goodwill G_{it} is specified such that it augments with marketing actions but decays over time and thus takes the geometric decay form

$$G_{it} = \beta' x_{it} + \lambda \beta' x_{i,t-1} + \lambda^2 \beta' x_{i,t-2} + \ldots = \sum_{j=0}^{\infty} \lambda^j \left(\beta' x_{i,t-j} \right), \qquad (2)$$

where x_{it} ($k_x \times 1$ dimension) represents time-varying independent variables that contribute to goodwill and β captures the corresponding marginal effects. The parameter λ is the carryover rate (correspondingly, $1-\lambda$ is the decay rate), which is assumed to be $0<\lambda<1$. Hence, the long-term effects of marketing mixes are captured using an infinite lag distribution. Combining **Equations (1)** and **(2)**, we arrive at a structured infinite distributed lag model:

$$y_{it} = \tilde{\alpha}_i + \sum_{j=0}^{\infty} \lambda^j \left(\beta' x_{i,t-j}\right) + \tilde{\gamma}' z_i + \nu_{it}\,. \tag{3}$$

A straightforward estimation of **Equation (3)** is infeasible, as doing so requires an infinite number of observations of the explanatory variables. In practice, one might seek to *approximate* **Equation (3)** by using P number of observable variables in a finite sample using the following form:

$$\begin{aligned} y_{it} &= \tilde{\alpha}_i + \sum_{j=0}^{P} \lambda^j \left(\beta' x_{i,t-j}\right) + \tilde{\gamma}' z_i + \eta_{it}, \\ \eta_{it} &= \sum_{j=P+1}^{\infty} \lambda^j \left(\beta' x_{i,t-j}\right) + \nu_{it}. \end{aligned}$$

However, observe that the associated new error term η_{it} is a combination of the original error term ν_{it}, and the approximation error $\sum_{j=P+1}^{\infty} \lambda^j \left(\beta' x_{i,t-j}\right)$, caused by omitting the unobserved x_{is}'s for periods $s \leq t-(P+1)$. This decomposition raises several concerns with using the above approximation form for estimation. First, an endogeneity problem arises due to the presence of lagged explanatory variables within the approximation error. As explanatory variables, such as advertising expenditures, are likely correlated across time, the orthogonality condition with regard to the error term η_{it} and explanatory variables $(x_{i,t},\ldots,x_{i,t\text{-}P})$ no longer holds. Second, the error terms η_{it} would exhibit high auto-correlation, induced by the recursive nature of the approximation errors, and would no longer be independently distributed. These concerns are likely to be aggravated when T, the total number of temporal observations, is small, which likely is the case in many real-world applications.

A more practical approach to estimation, while avoiding above concerns, is to substitute part of the infinite geometric sum in y_{it} by the discounted lagged dependent variable $\lambda y_{i,t\text{-}1}$ (Koyck, 1954), transforming **Equation (3)** into the geometric lag model

$$y_{it} = \lambda y_{i,t-1} + \beta' x_{it} + \gamma' z_i + u_{it}\,, \tag{4}$$

$$u_{it} = \alpha_i + \nu_{it} - \lambda \nu_{i,t-1}\,, \tag{5}$$

where $\alpha_i = \left(1-\lambda\right)\tilde{\alpha}_i$ and $\gamma = \left(1-\lambda\right)\tilde{\gamma}$.[4] As can be seen, the geometric lag model—a functional form expression of the N-A framework—in **Equations (4)** and **(5)** closely resembles the setting under a general dynamic panel data model.[5] The key difference is in the presence of the lagged idiosyncratic shock $\nu_{i,t\text{-}1}$, which adds to the existing dynamics that arise from unobserved heterogeneity α_i and the lagged dependent variable $y_{i,t\text{-}1}$. Each of these three components generates a unique pattern in the data. The unobserved heterogeneity α_i has a constant effect that persists over time. The lagged dependent variable $y_{i,t\text{-}1}$ (state dependence in the current empirical context) has a long-lasting yet diminishing effect over multiple time periods. In contrast, the lagged idiosyncratic shock $\nu_{i,t\text{-}1}$ has a negative effect at time t but fully disappears afterwards.

The structure of the unobservable term u_{it} in **Equation (5)** raises three major challenges in estimation: (i) controlling for unobserved heterogeneity α_i, (ii) addressing the endogeneity problem due to lagged dependent variable $y_{i,t\text{-}1}$ being correlated with the individual effect α_i within the error structure u_{is} for $s<t$, and (iii) addressing the serial correlation induced by the lagged idiosyncratic shock $\lambda\nu_{i,t\text{-}1}$.

The dynamic panel data methods of Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998) provide a practical approach that can tackle issues (i) and (ii) by first differencing and utilizing the lagged levels and lagged differences as instruments. Because of their practicality, dynamic panel data methods have been used extensively in the marketing and economics literature to examine phenomena of dynamic nature, including the N-A framework (Paton, 2002; Neumayer, 2004; Clark et al., 2009; Xiong and Bharadwaj, 2013; Terris-Prestholt and Windmeijer, 2016; Chung, 2017; Ye et al., 2017; Hirunyawipada and Xiong, 2018). However,

---

[4] The (Koyck) transformation is conducted by multiplying the carryover rate λ with the lagged form of **Equation (3)** and subtracting it from the current form.

[5] The dataset consists of (y_{i1}, y_{i2},..., y_{iT}), (x_{i1}, x_{i2},..., x_{iT}), and z_i for i=1,2,...,N, implying a dimension of $N\times T$ observations. The focus in the dynamic panel data analysis is mainly on the case where N is large and T is small, which is typical of the data that is available in real world studies (e.g., advertising).

these methods are only valid under the assumption of no serial correlation in the idiosyncratic errors, which violates (iii).

In the following subsections, we first outline the application of conventional dynamic panel data methods to serve as building blocks. Subsequently, we discuss the potential bias of dynamic panel data methods that arise due to the serial correlation present within the N-A framework and propose our methodology as a remedy that is robust to this bias.

2.2. Conventional GMM Estimation

The N-A framework represented by the geometric lag model in **Equations (4)** and **(5)** gives rise to the dynamic panel data methods (Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998).[6] While the conventional dynamic panel data methods postulate an identical regression equation to **Equation (4)**, the unobserved component in **Equation (5)**, which exhibits serial correlation, is substituted by

$$u_{it} = \alpha_i + \varepsilon_{it}\,, \tag{6}$$

where the idiosyncratic shock ε_{it} is assumed to be serially uncorrelated over time.

The challenge in estimating a dynamic panel data model is the endogeneity problem that arises from the time-invariant unobserved heterogeneity component α_i being correlated with the lagged dependent variable $y_{i,t\text{-}1}$. This issue can be dealt with in a relatively straightforward manner by taking the first difference of **Equation (4)** to subtract out α_i. However, the endogeneity problem with regard to the idiosyncratic error term (i.e., the lagged dependent variable $y_{i,t\text{-}1}$ being correlated with the lagged error terms ε_{is} for $s<t$) remains a concern. Hence, Anderson and Hsiao (1981, 1982), and Arellano and Bond (1991) utilize lagged dependent variables as instruments to derive the following moment conditions:

$$\mathrm{E}[y_{is}\Delta u_{it}] = 0 \tag{7}$$

[6] Formal statements of assumptions and details of algebraic derivations are stated in the Appendix. Interested readers are directed to Arellano (2003) for a more comprehensive discussion on the panel data methods.

for t=3,4,...,T and s=1,2,...,t−2, where $\Delta u_{it} = u_{it} - u_{i,t-1}$.[7] The estimator utilizing the moment conditions in **Equation (7)** is commonly referred to as the *Difference GMM* (DGMM) estimator. The DGMM uses the lagged variables in levels as instruments for the first differenced equation. However, a potential drawback of the DGMM estimator is that lagged levels become weak instruments for the first difference as λ becomes close to unity, where the lagged levels take a random walk and convey limited information (Staiger and Stock, 1997; Stock et al., 2002).

As a remedy, Arellano and Bover (1995) and Blundell and Bond (1998) propose utilizing both lagged differences and levels as instruments. Under their method, the following linear moment conditions become further available:

$$\mathrm{E}[u_{it}\Delta y_{i,t-1}] = 0 \tag{8}$$

for t=3,4,...,T. The estimator utilizing the moment conditions in both **Equations (7)** and **(8)** is commonly referred to as the *System GMM* (SGMM) estimator. The SGMM creates a stacked dataset that utilizes both lagged levels to instrument for differences (**Equation (7)**) and lagged differences to instrument for levels (**Equation (8)**). Thus, the SGMM estimator extracts more information from the data and benefits from an increased number of moment conditions.

The moment conditions pertinent to the explanatory variables x_{it} and z_i are derived from their relationship with the unobserved term u_{it}. Regarding the correlation between the explanatory variables and the idiosyncratic error ε_{it}, there are two commonly used assumptions: strict exogeneity and predetermined variables. The *strict exogeneity* assumption postulates that ε_{it} is uncorrelated with x_{is} for any s and z_i, which is very restrictive. Especially in marketing contexts, the time-varying explanatory variable x_{is} (e.g., advertisement expenditures) is often a choice of the firm. Hence, when choosing x_{is} at time s, the firm is likely to utilize past information collected from previous shocks ε_{it} for $t<s$, which would bring about correlation. The *predetermined variables* assumption relaxes the strict exogeneity assumption and allows the explanatory variables to be

[7] Following the standard notation in the literature, the capital Greek letter delta Δ represents a first-difference operator (e.g., $\Delta u_{it} = u_{it} - u_{i,t\text{-}1}$ and $\Delta y_{it} = y_{it} - y_{i,t\text{-}1}$).

correlated with past idiosyncratic shocks, but predetermined to current and future shocks. That is, explanatory variables are determined before current or future idiosyncratic shocks are realized; thus, the moment conditions are generated based on ε_{it} being uncorrelated with x_{is} for $s \leq t$ and z_i.[8]

We consider the less restrictive predetermined variables assumption as it is more plausible in various applications. In Section 5, we further relax this assumption so that explanatory variables are *weakly predetermined,* allowing for simultaneous realization of the current-period idiosyncratic shock (i.e., x_{is} is endogenous to the current-period shock and, hence, ε_{it} is uncorrelated with x_{is} for $s<t$ only), and discuss moment conditions to address this potential endogeneity.

Now, let us specify the relation between explanatory variables and the individual effect α_i. We allow a section of time-varying explanatory variables to be correlated with the individual effect α_i, and partition the vector as $x_{it}=(x_{1it},x_{2it})$. Here, x_{1it} is a vector orthogonal to the individual effect, whereas x_{2it} is not—that is, the latter is correlated with the individual effect (Hausman and Taylor, 1981). Then, in addition to **Equations (7)-(8)**, the following moment conditions become available[9]:

$$\begin{aligned}
&\mathrm{E}[x_{1is}u_{it}] = 0 \quad \text{for } s \leq t, \\
&\mathrm{E}[x_{2is}\Delta u_{it}] = 0 \quad \text{for } s \leq t-1, \\
&\mathrm{E}[u_{it}\Delta x_{2i,t}] = 0 \quad \text{for } t = 2,3,...,T.
\end{aligned} \tag{9}$$

To identify γ, the marginal effect with regard to the time-unvarying variable z_i, using SGMM[10], we use the *random effect model* for the individual effect α_i. This requires that z_i be orthogonal to the individual effect α_i (in addition to z_i's predetermined nature). If not, the effect of z_i is absorbed

[8] Utilizing predetermined variables in panel data models is discussed in Hausman and Taylor (1981), Amemiya and MaCurdy (1986), and Breusch et al. (1989).

[9] In DGMM, the conditions in **Equation (9)** reduce to $\mathrm{E}[x_{is}\Delta u_{it}]=0$ for $s \leq t-1$. As DGMM takes only the first-differencing of the equation into account, the partitioning of the vector $x_{it}=(x_{1it},x_{2it})$ relative to individual effects α_i no longer plays a role in determining the moment conditions.

[10] In line with footnote 9, the parameter γ cannot be identified using DGMM due to first-differencing of the equation.

into the individual effect α_i and γ cannot be separately identified. Given this assumption, in addition to **Equations (7)-(8)**, the following moment conditions are available for estimation[11]:

$$\mathrm{E}[z_i u_{it}] = 0 \quad \text{for } t = 1, 2, ..., T\,. \tag{10}$$

The key advantage of the above dynamic panel data methods is that they do not rely on strictly exogenous instruments. As valid exogenous instruments are often difficult to come by, dynamic panel data methods have been widely applied in various studies across diverse topics (Clark et al., 2009; Song et al., 2015; McAlister et al., 2016; Van Triest et al., 2009; Tuli et al., 2010; Rego et al., 2013; Narasimhan et al., 2006; Fang et al., 2016; Shah et al., 2014; Narayan and Kadiyali, 2016; Mathys et al., 2016; Chung, 2017; Archak et al., 2011; Germann et al., 2015; Feng et al., 2015; Mukherji et al., 2011; Burtch et al., 2013; Acemoglu et al., 2008; Durlauf et al., 2005).

The underlying micro-foundation—that is, the assumption on the data generating process (DGP)—of past studies that have used the dynamic panel data methods largely falls into two categories. First, studies have used the lagged dependent variable simply as a control with no direct construct of the underlying DGP (i.e., no interpretation of the micro-foundation). Second, studies, especially in advertising, have used the N-A framework, either explicitly or implicitly, as the micro-foundation behind the dynamics (Paton, 2002; Neumayer, 2004; Clark et al., 2009; Xiong and Bharadwaj, 2013; Terris-Prestholt and Windmeijer, 2016; Chung, 2017; Ye et al., 2017; Hirunyawipada and Xiong, 2018).

In terms of the model and estimation, past studies have used the empirical approach represented in **Equation (4)** with the unobserved component in **Equation (6)**. However, as shown from **Equations (1)-(5)**, serial correlation in the error structure exists by construction under the micro-foundation (and thus the DGP) of the N-A framework. Hence, if one does not account for serial correlation, the resulting estimates become biased, leading to incorrect inference. We elaborate on the direction of these biases in the next subsection.

[11] Although having z_i that is not orthogonal to α_i does not identify γ, it can aid efficient estimation of other parameters of the model (i.e., λ, β) through the use of the following moment conditions: $\mathrm{E}[z_i \Delta u_{it}]=0$ for t=2,3,..., T.

2.3. Bias Associated with Serial Correlation

The conventional GMM estimators are biased when the DGP follows the N-A framework. Under the framework, $\mathrm{E}[y_{i,t-1}u_{it}] \neq 0$ and $\mathrm{E}[x_{it}u_{it}] \neq 0$ as u_{it} includes $\nu_{i,t\text{-}1}$, which is correlated with $y_{i,t\text{-}1}$ and x_{it}. More specifically, the following moment conditions become invalid:

$$
\begin{aligned}
&\mathrm{E}[y_{i,t-2}\Delta u_{it}] = 0,\\
&\mathrm{E}[u_{it}\Delta y_{i,t-1}] = 0,\\
&\mathrm{E}[x_{1it}u_{it}] = 0,\\
&\mathrm{E}[x_{2i,t-1}\Delta u_{it}] = 0,\\
&\mathrm{E}[u_{it}\Delta x_{2it}] = 0.
\end{aligned}
$$

Prior to formal examination of the biases associated with each of the invalid conditions above, let us illustrate the course of how the biases arise. The key task in dynamic panel data methods is to explain the serial correlation in the dependent variable (i.e., $\mathrm{E}[y_{it}y_{i,t-1}]$). By replacing y_{it} with $\lambda y_{i,t-1} + \beta' x_{it} + \gamma' z_i + u_{it}$, we can decompose the serial correlation into four terms:

$$
\mathrm{E}[y_{it}y_{i,t-1}] = \underbrace{\lambda \mathrm{E}[y_{i,t-1}^2] + \beta' \mathrm{E}[x_{it}y_{i,t-1}] + \gamma' \mathrm{E}[z_i y_{i,t-1}]}_{observed} + \underbrace{\mathrm{E}[u_{it}y_{i,t-1}]}_{unobserved}.
$$

In this decomposition, the first three terms are determined by the parameters and expectations based on observed variables, but the last term includes an unobserved variable. Conventional dynamic panel data methods remove the unobserved term by taking the first-difference—which is made possible as u_{it} is assumed to consist of only the time-invariant individual effect and the idiosyncratic error uncorrelated with $y_{i,t\text{-}1}$. However, under the N-A framework, the unobserved term cannot be fully removed by first-differencing:

$$
\mathrm{E}[u_{it}y_{i,t-1}] = \mathrm{E}[\alpha_i y_{i,t-1}] + \mathrm{E}[\nu_{it}y_{i,t-1}] - \lambda \mathrm{E}[\nu_{i,t-1}y_{i,t-1}]
$$

as the rightmost term is neither cancelled out by first-differencing nor accommodated in the model. Thus, when conventional methods are fitted to the data under the N-A framework, this unaccommodated term becomes absorbed into other terms. Because the sign of the unaccommodated term is negative ($-\lambda \mathrm{E}[\nu_{i,t-1}y_{i,t-1}] = -\lambda \mathrm{E}[\nu_{i,t-1}^2] < 0$), the other terms are likely to be underestimated. Based on this observation, we can infer the direction of the biases. For example, because $\lambda \mathrm{E}[y_{i,t-1}^2]$ is expected to be underestimated and $\mathrm{E}[y_{i,t-1}^2] > 0$, λ is also expected to be underestimated; thus, it becomes downward biased.

Now, let us examine the direction of biases with more rigor. For brevity, we outline an illustrative case in the main script (a complete set of derivations is provided in the Appendix). A moment condition is invalid if it is different from zero when evaluated at the true parameter. The solution to an invalid moment condition is a (not true) parameter value that makes the moment condition equal to zero, so the estimator is biased towards the solution. Thus, the direction of the bias can be determined by using the sign of the moment condition at the true parameter value and the slope of the moment condition with respect to the parameter.

For example, we can observe that the invalid moment condition $\mathrm{E}[y_{i,t-2}\Delta u_{it}] = 0$ is positive at the true parameter because $\mathrm{E}[y_{i,t-2}\Delta u_{it}] = \lambda \mathrm{E}[\nu_{i,t-2}^2] > 0$. By replacing u_{it} in the moment condition with $y_{it} - \lambda y_{i,t-1} - \beta' x_{it} - \gamma' z_i$, we can decompose $\mathrm{E}[y_{i,t-2}\Delta u_{it}]$ as:

$$\mathrm{E}[y_{i,t-2}\Delta u_{it}] = \mathrm{E}[y_{i,t-2}\Delta y_{it}] - \lambda \mathrm{E}[y_{i,t-2}\Delta y_{i,t-1}] - \beta' \mathrm{E}[y_{i,t-2}\Delta x_{it}],$$

which is a linear function of the parameters. This is increasing in λ because $-\mathrm{E}[y_{i,t-2}\Delta y_{i,t-1}] > 0$. Because the moment condition is positive at the true parameter and its slope with respect to λ is positive, the solution to the moment condition is smaller than the true parameter. Therefore, it follows that the invalid moment condition $\mathrm{E}[y_{i,t-2}\Delta u_{it}] = 0$ leads to a downward bias in λ.

Table 1 summarizes the full set of biases resulting from invalid moment conditions. Although the effects of invalid moment conditions are mixed for the parameter λ, we expect the conventional estimator to be downward biased because the first two moment conditions are typically more informative than the latter. This conjecture is also consistent with our simulation results shown in Section 4. All invalid moment conditions also cause a downward bias in β_1 and β_2. However, the bias in γ remains uncertain as the variable z_i is usually cancelled out by first-differencing.

2.4. Restricting the Moment Conditions

The remedy for the misspecification bias is not to use the invalid moment conditions with regard to the lagged dependent variable in the estimation (Hujer et al., 2005; Lee et al., 2017). In addition, some of the moment conditions regarding the explanatory variable become invalid. While past studies have focused on the endogeneity issue with regard to the lagged dependent variable, we also deal with the endogeneity issue related to the predetermined variables.

We propose a restricted set of moment conditions that are immune to serial correlation in the error structure. By removing the invalid moment conditions from **Equations (7)** and **(8)**, the following set of conditions remain valid for DGMM:

$$\mathrm{E}[y_{is}\Delta u_{it}] = 0 \tag{11}$$

for t=4,5,...,T and s=1,2,...,t−3, and additionally for SGMM:

$$\mathrm{E}[u_{it}\Delta y_{i,t-2}] = 0 \tag{12}$$

for t=4,5,...,T.

By accounting for serial correlation in the error structure with regard to time-varying explanatory variables x_{it}, in addition to **Equations (11)** and **(12)**, the following moment conditions remain valid[12]:

$$\begin{aligned} &\mathrm{E}[x_{1is}u_{it}] = 0 \quad \text{for } s \leq t-1 \\ &\mathrm{E}[x_{2is}\Delta u_{it}] = 0 \quad \text{for } s \leq t-2 \\ &\mathrm{E}[u_{it}\Delta x_{2i,t-1}] = 0 \quad \text{for } t = 3,4,...,T\,. \end{aligned} \tag{13}$$

However, the moment conditions in **Equation (10)** regarding time-constant regressors z_i are not directly affected by serial correlation.

In the subsequent analysis, we compare the performance of three types of GMM estimators using different sets of the moment conditions (see specifications in **Table 2**). We refer to the GMM estimator using the following conditions: (i) conventional moment conditions for both the lagged dependent and predetermined variables (**Equations (7)-(10)**) as the *conventional* estimator; (ii) restricted moment conditions for the lagged dependent variables but conventional moment conditions for the predetermined variables (**Equations (9)-(12)**) as the *restricted*

[12] Analogous to footnote 9, in DGMM, the conditions in **Equation (13)** reduce to E[$x_{is}\Delta u_{it}$]=0 for $s \leq t-2$.

estimator; and (iii) restricted moment conditions for both the lagged dependent and predetermined variables (**Equations (10)-(13)**) as the *restricted-pre* estimator.[13]

The conventional estimator becomes biased under the N-A framework due to serial correlation in the error structure. The validity of the restricted and the restricted-pre estimators hinges on the nature of the time-varying explanatory variables x_{is}. If x_{is} is strictly exogenous (i.e., x_{is} is uncorrelated with ν_{it} for any t=1,2,...,T), both the restricted estimator and the restricted-pre estimator become unbiased, but the restricted estimator is more efficient than the restricted-pre estimator. However, if x_{is} is predetermined (i.e., uncorrelated with ν_{it} for $t \geq s$ only), the restricted-pre estimator is the only unbiased estimator.

3. Test for Serial Correlation

As discussed in the previous section, overlooking serial correlation in the error structure can result in biased estimates that lead to incorrect inference. The Arellano and Bond (1991) specification test—specifically, the AR(2) test—has been widely employed in empirical applications to check for serial correlation in idiosyncratic errors, and therefore the validity of the model specification. The AR(2) test checks for the second-order serial correlation in the error structure differences, and thus the first-order serial correlation in the levels. If the AR(2) test is not rejected (i.e., the error structure is presumably serially uncorrelated), studies have adopted the use of conventional moment conditions in **Equations (7)-(9)** without further caution.

The AR(2) test, however, is prone to weak power and often fails to reject the null hypothesis. That is, the AR(2) test may indicate the absence of serial correlation even under its presence, and thus would wrongly justify the use of conventional moment conditions—part of which are invalid. This could lead to unreliable and biased inference. Studies have found the AR(2) test to exhibit weak power if (i) the test statistic is constructed using biased estimates obtained from invalid moment conditions (Jung, 2005), or (ii) too many moment conditions are employed under finite samples (Bowsher, 2002).

[13] The moment conditions in **Equation (10)** apply analogously to all cases.

Let us illustrate our proof on how the AR(2) test may fail to reject the null hypothesis (of no serial correlation) in models under the N-A framework. Specifically, we consider the case where the DGP follows the N-A framework, but the test statistic is built upon using conventional moment conditions—as in previous studies that apply the dynamic panel data methods.

Suppose the true model is represented by the geometric lag model in **Equations (4)** and **(5)**, of which the error structure is serially correlated. However, without taking into account the possibility of serial correlation, suppose a researcher estimates the carryover parameter λ using the conventional moment conditions given in **Equations (7)-(9)**. Denote the resulting estimate by $\hat{\lambda}$. As discussed in the previous section, the conventional estimator $\hat{\lambda}$ is expected to be downward-biased. For brevity, let $B = \mathrm{E}(\hat{\lambda} - \lambda)$ denote this bias.

In the AR(2) test, we are particularly interested in testing the null hypothesis of no serial correlation $\mathrm{E}[\Delta u_{it}\Delta u_{i,t-2}] = 0$ against its negation. Let $\widehat{\Delta u_{it}} = \Delta y_{it} - \hat{\lambda}\Delta y_{i,t-1}$ be the sample estimate of Δu_{it}. Due to the bias in $\hat{\lambda}$, the estimate $\widehat{\Delta u_{it}} \approx \Delta u_{it} - B\Delta y_{i,t-1}$ also becomes contaminated. By replacing the expectation with the sample counterparts, we have:

$$\mathrm{E}[\widehat{\Delta u_{it}}\widehat{\Delta u_{i,t-2}}] \approx \mathrm{E}[\Delta u_{it}\Delta u_{i,t-2}] - B\mathrm{E}[\Delta y_{i,t-1}\Delta u_{i,t-2}] - B\mathrm{E}[\Delta y_{i,t-3}\Delta u_{it}] + B^2\mathrm{E}[\Delta y_{i,t-1}\Delta y_{i,t-3}].$$

For illustrative purposes, let us assume homoskedasticity in differences (i.e., $E(\nu_{it}^2) = \sigma_\nu^2$ for all t). By substituting the components in **Equations (4)** and **(5)**, the above terms become:

$$\begin{aligned}
&\mathrm{E}[\Delta u_{it}\Delta u_{i,t-2}] = \lambda\sigma_\nu^2 > 0,\\
&\mathrm{E}[\Delta y_{i,t-1}\Delta u_{i,t-2}] = -\sigma_\nu^2,\\
&\mathrm{E}[\Delta y_{i,t-3}\Delta u_{it}] = \mathrm{E}[\Delta y_{i,t-1}\Delta y_{i,t-3}] = 0.
\end{aligned}$$

Note that, in the absence of bias (B=0), $\mathrm{E}[\widehat{\Delta u_{it}}\widehat{\Delta u_{i,t-2}}]$ converges to $\lambda\sigma_\nu^2$ and the degree of serial correlation captured by λ and σ_ν^2 jointly determines the test statistic. However, when the biased estimate $\hat{\lambda}$ is used, $\mathrm{E}[\widehat{\Delta u_{it}}\widehat{\Delta u_{i,t-2}}]$ converges to $(\lambda + B)\sigma_\nu^2$. Because B is negative, $\mathrm{E}[\widehat{\Delta u_{it}}\widehat{\Delta u_{i,t-2}}]$ also becomes downward biased. Hence, depending on the degree of bias in $\hat{\lambda}$, the test statistic based on $\mathrm{E}[\widehat{\Delta u_{it}}\widehat{\Delta u_{i,t-2}}]$ may falsely infer that $\mathrm{E}[\widehat{\Delta u_{it}}\widehat{\Delta u_{i,t-2}}] = 0$, and fail to reject the null hypothesis, indicating a lack of serial correlation.

Bowsher (2002) argues that the bias described above and the resulting weak power of the AR(2) test is particularly problematic with finite samples. Although the AR(2) test statistic could be biased under the alternative hypothesis, asymptotically this bias would not affect the performance of the test—as the standard error converges to zero, the test statistic becomes precise enough to overcome the bias. However, in a finite samples setting with large standard errors, the bias becomes costly when undetected.

We suggest the use of test statistics composed from the restricted or the restricted-pre estimators to check for serial correlation. These estimators remain unbiased (even with finite samples) and do not suffer from the problem caused by the downward bias in λ. Consequently, if the test is rejected, moment conditions in **Equations (10)-(13)** (**Equations (9)-(12)** for restricted) should be used for estimation. However, if the test is not rejected, which would give more certainty of no serial correlation than the conventional tests, moment conditions in **Equations (7)-(10)** could be used to fully exploit all valid information to yield greater efficiency.

The intuition behind our proposed method of testing is similar to Jung (2005), who proposes using consistent estimators to obtain the test statistic under a more general serial correlation structure such as autoregressive (AR) or the moving-average (MA) model of higher orders. However, our approach differs in at least three dimensions: (i) we consider the time-varying explanatory variables, which also cause inconsistency; (ii) we use micro-foundation of the N-A framework, which provides a priori knowledge on the form of serial correlation; and (iii) the N-A framework generates a negative serial correlation, which is not covered in Jung (2005).

In the next section, we verify the above assertions using simulation studies and thus demonstrate the following: poor performance of the test statistic under conventional methods when serial correlation is present; and the test based on the restricted and the restricted-pre estimators yield better power properties in finite samples.

4. Simulation Study

To compare and evaluate the performance of different estimators and the Arellano-Bond specification test, we conduct Monte Carlo experiments using simulated data. The DGP is set to follow the N-A framework with one predetermined variable (k_x=1) such that

$$y_{it} = \tilde{\alpha}_i + \sum_{j=0}^{\infty} \lambda^j \left(\beta x_{i,t-j}\right) + \nu_{it} \,.$$

We allow for heteroskedasticity in the idiosyncratic shock, namely $\nu_{it} \overset{\text{i.i.d.}}{\sim} N(0,\sigma_\nu^2)$, where $\sigma_{\nu,it}^2 = \theta_0 + \theta_1 \cdot x_{it}^2$, and let the individual effect $\tilde{\alpha}_i = \left(\frac{1}{1-\lambda}\right)\alpha_i$, where $\alpha_i \overset{\text{i.i.d.}}{\sim} U(0,1)$.[14]

Regarding the time-varying explanatory variable x_{it}, the following equation is considered

$$x_{it} = \rho x_{i,t-1} + \varphi\tilde{\alpha}_i + \tau\nu_{i,t-1} + \xi_{it} \,,$$

where $\xi_{it} \overset{\text{i.i.d.}}{\sim} N(0,\sigma_\xi^2)$. Hence, x_{it} is specified such that it follows an AR(1) process, exhibits correlation with individual effects, and is allowed to incorporate past shocks ν_{is} for $s<t$ (i.e., predetermined with respect to ν_{is} for $t \leq s$). The parameter values are set as θ_0=0.8, θ_1=0.2, $\sigma_\xi = 2$, and ρ=φ=0.3, and the data is generated for N=500 and T=8—a typical structure of dynamic panel data where N is large and T is small.[15,16] We run 200 Monte Carlo iterations and report the mean values and standard deviations of the estimates.

In the following, the robustness of *conventional*, *restricted*, and *restricted-pre* estimators is considered under two different scenarios; whether x_{it} is predetermined (τ=0.5) or strictly exogenous (τ=0) with respect to past shocks ν_{is} for $s<t$.

4.1. Case 1: Predetermined Explanatory Variable (τ=0.5)

[14] While α_i does not necessarily affect the estimation due to first-differencing, it influences the degree of correlation across variables, and hence, the above transformation helps maintain consistency of the simulation across different λ values.

[15] We follow Arellano and Bond (1991) in the base design of the exercise. We extend their settings to allow x_{it} to be predetermined and correlated with the individual effects, and assume the DGP to follow the N-A framework.

[16] To incorporate the infinite lag structure of the goodwill formulation, we draw 500 prior observations of $x_{i,t-j}$ (j=1,2,...,500) per individual to formulate the initial (t=0) goodwill. For λ=0.9, $\lambda^{500} \approx 1.32 \times 10^{-23}$.

When transformed into a dynamic panel data methods setting, the above DGP exhibits serial correlation in the errors as demonstrated in **Equations (3)-(5)**. Hence, the conventional estimator is likely biased in the direction illustrated in Section 2.3 (**Table 1**). The restricted estimator, while correcting for the biases from using the lagged dependent variable, utilizes invalid moment conditions for predetermined explanatory variables x_{it}, and likely leads to the biases listed in the bottom rows of **Table 1**. By correcting for invalid moment conditions with respect to both the lagged dependent variable and predetermined variable, the restricted-pre estimator is expected to be robust.

The simulation results of DGMM and SGMM under the three estimators are reported in **Table 3**. The upper portion of each table presents the mean and standard deviations of the estimates across iterations; the lower portion reports the rejection frequency of the Arellano-Bond specification test.

Table 3a reports the estimation results of the conventional estimator. As can be seen, the conventional methods (in both DGMM and SGMM) exhibit strong bias. Consistent with our theoretical analyses in Section 2.3, both λ and β estimates fail to recover their true values and are downward biased. The magnitude intensifies as the carryover rate λ increases because the invalid instruments become strong, conveying more unreliable information.

In **Table 3b**, the λ estimates under the restricted method show significant recovery. After correcting for the misspecification with regard to the lagged dependent variables, the remaining bias tends upward for the λ parameter (see **Table 1**, bottom two rows). This is verified across different values of λ. In addition, the restricted estimator suffers from the invalid moment conditions regarding the predetermined variables. This is represented by the downward-biased β estimates, especially as λ goes to unity.

The simulation results for the restricted-pre estimator appear in **Table 3c**. The mean estimates indicate that the method recovers the model primitives well. A slight exception occurs for DGMM as λ approaches unity, where the carryover-rate estimates become downward-biased.

This results from the aforementioned weak-instruments problem, where the lagged levels lose information as $\lambda \approx 1$ and become poor instruments for the first-differences (Blundell and Bond, 1998). The SGMM estimates remain robust across all parameter values.

Now, let us turn our attention to the test statistics. The null hypothesis for the Arellano-Bond AR(1) and AR(2) tests is that there exists *no* first- and second-order serial correlation, respectively, in the differenced error structure. The results across the three estimators show that AR(1) tests are rejected almost all of the time, implying that a first-order serial correlation exists among the differences. This is to be expected, as Δu_{it} and $\Delta u_{i,t-1}$ are correlated through the shared $\nu_{i,t\text{-}1}$ term by construction.

The AR(2) test checks for the existence of second-order serial correlation in differences and thus first-order serial correlation in levels. This information is used to rationalize the use of conventional moment conditions—that is, if the AR(2) test is not rejected, serial correlation in levels is not expected and thus the use of conventional moment conditions is justified. The following patterns are noteworthy from the AR(2) test results of the simulation exercise.

First, the test based on the restricted and the restricted-pre estimators correctly rejects the null from the use of unbiased λ estimates; in contrast, the power of the test using the conventional approach decreases due to biased estimates, as discussed in Section 3. For example, in **Table 3a**, the test based on the conventional approach in SGMM when λ=0.8, rejects the null in only 63.5% of the iterations. This implies that for the remaining 36.5%, the underlying serial correlation—present in the case of the N-A framework or any other framework following the DGP of the geometric lag model—would go undetected.

Second, regardless of the approach, the power of the test based on DGMM decreases significantly as λ goes to unity, due to the weak-instrument problem discussed in Section 2. In contrast, the test based on SGMM remains robust from the weak-instrument problem, where rejection frequency remains steady over different values of λ.[17]

4.2. Case 2: Exogenous Explanatory Variable (τ=0)

Now, let us consider the case where x_{it} is strictly exogenous (with respect to past shocks ν_{is} for $s<t$), and is no longer predetermined. Although the aforementioned DGP leads to serial correlation in the error structure, the moment conditions pertaining to the time-varying explanatory variables in **Equation (9)** now hold. Hence, the moment conditions utilized by the restricted and the restricted-pre estimators are both valid—with the restricted estimator utilizing more conditions.

The results of this series of simulation experiments are reported in **Table 4**. As can be seen from **Tables 4b** and **4c**, both the restricted and the restricted-pre estimators perform adequately at recovering the parameter values. The restricted estimator, however, shows greater efficiency (lower variance) from the use of more (valid) moment conditions. This efficiency gain also carries over to the test statistics, and overall, the restricted estimator yields slightly better power properties than the restricted-pre counterpart.

The conventional estimator in **Table 4a,** however**,** fails to recover the true parameter values. While the validity of moment conditions regarding the explanatory variables reduce the degree of downward bias in the β estimates, in general, both λ and β estimates remain downward biased. A notable difference compared to the above predetermined case is the significant decrease in power of the AR(2) test using the conventional estimator. Because invalid moment conditions that generated the upward bias in λ no longer exist, what remains for λ is even greater downward bias. As a result, now the test statistics reject the null, when λ=0.8, for only 45.5% of the iterations.

The key findings of the two simulation exercises are summarized as follows: when the underlying DGP follows the N-A framework, the following occurs: (i) the conventional estimator

[17] The low power of the test across all approaches for λ=0.1 to 0.2 reflects the actual limited degree of serial correlation.

fails to recover the true parameters and becomes downward biased; (ii) if the explanatory variable is strictly exogenous, both the restricted and the restricted-pre estimators are robust, with the restricted estimator being more efficient; and (iii) if the explanatory variable is predetermined, only the restricted-pre estimator remains robust. Thus, we suggest the use of moment conditions of the restricted-pre estimator if strict exogeneity of the explanatory variables is in question. In addition, the restricted-pre estimator could serve as a device to test for the validity of the strict exogeneity assumption for which the restricted approach yields more efficient estimates.

Furthermore, the simulation results verify our claims on testing for serial correlation. Under the N-A framework, the AR(2) test statistics obtained by conventional dynamic panel data methods fail to adequately detect serial correlation in the error structure, wrongly justifying the use of the biased estimates. In practice, this bias may lead to misinterpreting or undermining the true marginal effects of the covariates. In contrast, the restricted and restricted-pre approaches are robust in detecting plausible serial correlation.

5. Empirical Analysis

In this section, we apply our method to real-world data to properly examine the short- and long-term effects of detailing efforts. First, we describe the institutional setting and data and then present the empirical model. We discuss the results in Section 6.

5.1. Data and Institutional Details

The focal firm is a highly regarded Fortune 500 company that operates in over 150 countries. It offers a broad range of branded generic pharmaceuticals, along with medical devices, diagnostics, and nutritional products. Our empirical analysis utilizes data from the chronic-care sales division of the firm's business operations in India. The data consist of a detailed record of prescriptions written by physicians over a six-month period from January through June 2016. For each physician, we observe the number of prescriptions written and the number of visits by the sales representatives.

The firm organizes its sales activity by route call sales. At the beginning of each month, the regional manager, together with the sales representative, creates a route plan that includes a series of scheduled visits and brands to be detailed to each physician. During the month, the sales rep makes in-person sales calls following the assigned route and updates the physician detailing report—e.g., number of prescriptions and special campaigns. The firm's compensation plan for the sales representatives is based on salary and commissions, where the latter is incentivized purely based on the sales performance outcomes.

Our data are unique in that they include the full range of the firm's brands. Previous studies of detailing effectiveness (Parsons and Vanden Abeele, 1981; Manchanda et al., 2004; Mizik and Jacobson, 2004) have limited their data to a single brand or a few brands. Consequently, their results neglect possible spillover effects among brands, and thus can potentially underestimate the overall effectiveness of detailing. Because firms and managers are likely to be most interested in the impact of sales calls on overall performance, we believe our dataset provides an appropriate measure to evaluate the effect of detailing efforts on sales.[18]

To fully exploit the nature of a dynamic panel data model, we restrict our attention to physicians with ongoing salesforce interactions, and for whom the data include no intermissions in prescription history. To explore differences in the effectiveness of sales calls across physician specialties, we focus exclusively on the six medical practice areas that account for approximately 90% of the active physicians in our data: cardiologists, diabetologists, endocrinologists, consulting physicians, general practitioners, and general surgeons. For expository purposes, we refer to the first three groups as *specialists* and the latter three as *generalists*. These restrictions lead us to focus our attention on N=9,595 physicians over T=6 month horizon.

Figure 1 depicts the empirical distribution of prescriptions and detailing calls. **Figure 1a** shows the number of prescriptions per month to be highly heterogeneous and right-skewed across doctors, implying significant unobserved physician heterogeneity. The number of calls per month, illustrated in **Figure 1b**, shows heterogeneity similar in shape to **Figure 1a** but also discreteness:

[18] Because the firm does not track the specific brands detailed during each call, we aggregate prescription quantities across the firm's brands to obtain the total quantity of prescriptions written per month. The aggregation of prescriptions is valid in the current context of generics pharmaceutical products, where prices are relatively homogeneous and contained within $1-3 range.

the majority of observations fall between 1 and 7 visits, in keeping with the firm's route sales procedure.

Table 5 reports descriptive statistics by medical practice area for the number of prescriptions and detailing calls. Most striking is the magnitude of the between-group difference: both sales-force efforts and outcomes are, on average, greater for the specialists. Heterogeneity in both prescriptions and sales calls within the specialists group is also noteworthy: although the diabetologists write the most prescriptions, sales-force efforts are more intensively targeted at the endocrinologists. Among generalists, however, within-group heterogeneity is only modest: consulting physicians generate slightly more attention and sales.

5.2. The Empirical Model

We model doctor i's prescriptions of the focal firm's pharmaceutical drugs at time t, S_{it}, as a function of an unobserved doctor-specific effect $\tilde{\alpha}_i$ constant over time, a stock of goodwill G_{it} (created by the firm's sales force), a time-specific effect (reflecting seasonality) $\tilde{\delta}_t$ common to all physicians, and an idiosyncratic unobserved component ν_{it} such that:

$$S_{it} = \exp\left(\tilde{\alpha}_i + G_{it} + \tilde{\delta}_t + \nu_{it}\right).$$

The multiplicative form is used to prevent overweighting of high-volume prescribers. Assuming the stock of goodwill G_{it} follows the N-A framework and thus the geometric decay form in **Equation (2)**, the transformation illustrated in **Equations (1)-(5)** is exploited to simplify the model specification into

$$\begin{aligned} s_{it} &= \lambda s_{i,t-1} + \beta' x_{it} + \delta_t + u_{it} \qquad (14) \\ u_{it} &= \alpha_i + \nu_{it} - \lambda \nu_{i,t-1} \end{aligned}$$

where $s_{it} = \log(S_{it})$, $\alpha_i = (1-\lambda)\tilde{\alpha}_i$, and $\delta_t = \tilde{\delta}_t - \lambda\tilde{\delta}_{t-1}$.[19,20,21] Our empirical application uses the total number of detailing calls provided to physician i during month t for x_{it}. Notice again that by

[19] Because our model is multiplicative, the carryover parameter λ represents elasticity in the current setting.

[20] Regarding the initial condition of the data, we assume mean stationarity in the relationship between physicians and the focal firm. Algebraic statements and implications of this assumption are provided in the Appendix.

the geometric sum assumption of the stock of goodwill, the error structure, by construction, exhibits serial correlation.

The identification of detailing effectiveness relies on the variation in frequency of sales calls within a physician over time. **Figure 2** shows the distribution of this change (i.e., $\Delta x_{it}=x_{it}-x_{i,t-1}$) in the number of calls within a physician. We can see that there exists sufficient variation over time. This variation indicates that the firm strategically adjusts the level of detailing based on past outcomes of its sales efforts. However, this would imply that there may also be an endogeneity problem. Recall our assumption regarding predetermined variables: current-period idiosyncratic shocks are uncorrelated with current and lagged values but not necessarily with future values (ν_{it} is uncorrelated with x_{is} for $s \leq t$).

The idiosyncratic shocks (both ν_{it} and ν_{it-1}) in **Equation (14)** represent any factors not observed to a researcher but that affect the number of physician prescriptions—for example, a clinic closes temporarily for office renovation or staff vacations. If such events are planned to take place in a certain month and, knowing this, the firm adjusts call patterns for that month, then the number of calls would be endogenous. Hence, we relax our assumption on the explanatory variable from predetermined to *weakly predetermined*. That is, we allow current-period detailing efforts to be correlated with both past *and current* period idiosyncratic shocks, but predetermined with regard to future shocks. Hence, the idiosyncratic shocks are now uncorrelated only with the lagged values of the detailing efforts, but not necessarily with the current and future values (i.e., ν_{it} is uncorrelated with x_{is} for $s<t$). This implies that moment conditions in **Equation (13)** become valid one-lag below.

For the empirical estimation, we construct moment conditions for **Equation (14)** under three different assumptions with regard to serial correlation, as discussed in previous sections: conventional, restricted, and restricted-pre estimators (see **Table 2**). The difference from the

[21] We tested for diminishing returns of detailing efforts by including quadratic terms, as in Manchanda and Chintagunta (2004). However, all coefficients for the quadratic terms are found to be insignificant. Thus, we exclude them from the analysis.

specification in **Table 2** is that for the restricted-pre estimator, we relax our assumption on the explanatory variables from predetermined to weakly predetermined to address the potential endogeneity concern with regard to detailing efforts. We limit the total number of moment conditions for our predetermined variable (number of detailing calls) by using only the most recent lag available for the differenced equation to prevent a potential overfitting problem.

6. Results

We first discuss the results of the homogeneous model and then the heterogeneous model with different detailing effectiveness by physician specialty. Subsequently, we show that traditional tests may fail to detect the presence of serial correlation, resulting in a misuse of moment conditions for estimation. Finally, we discuss the managerial implications in optimizing call patterns for route sales.

6.1. Homogeneous Model

Table 6 reports the parameter estimates of the model given in **Equation (14)**. We first turn our attention to the specification test results. The Arellano-Bond test for serial correlation shows that both AR(1) and AR(2) are rejected across all specifications. This result implies the existence of both first- and second-order serial correlation in the differenced error structure, providing a strong rationale for restricted use of the instruments for the lagged dependent variable.

Because serial correlation exists in the unobserved components of the data, the key assumption under conventional methods is not satisfied. Hence, the estimates obtained using the improper moment conditions of conventional methods are biased, as shown in the first and second columns of **Table 6**. This is evident in the counterintuitive results, which indicates negative or minimal effectiveness of detailing. The carryover elasticity estimates also become downward-biased, as discussed in Section 2.3 and demonstrated in the Monte Carlo simulation.

By correcting for the invalid moment conditions of the lagged dependent variables, the restricted methods in the third and fourth columns show recovery in the carryover elasticity estimates. However, due to misspecifications regarding the predetermined nature of detailing calls, the slope parameters remain downward-biased, indicating limited effectiveness of detailing.

In the fifth column of **Table 6**, the DGMM estimates under the restricted-pre method also show recovered carryover elasticity measures from valid moment conditions. However, the model suffers from the weak instruments problem associated with the sole use of levels as instruments for differences, and the slope parameters representing the effectiveness of sales efforts remain insignificant. Thus, for model inference, we turn our attention to the SGMM estimates under the restricted-pre approach, which impose proper moment conditions while extracting more information from the data to correct for the weak instrument problem. In the far right column, we find that, on aggregate, the long-term effect—specifically, the carryover effect—is 0.562, and that in the short-term, a unit increase in detailing calls elicits an 11.85% increase in prescriptions by the physician.[22]

6.2. Heterogeneity in Detailing Effectiveness across Specialties

The preceding section accounts only for permanent heterogeneity using physician fixed effects. In reality, firms care about the effectiveness of detailing across different medical specialties. To investigate differences in the value of sales efforts across specialties, we allow for different slope parameters for each specialty such that

$$s_{it} = \sum_{d} I_{(i \in S_d)}(\lambda_d s_{i,t-1} + \beta_d{}' x_{it}) + \delta_t + u_{it} \tag{15}$$

where $I_{(i \in S_d)}$ is an indicator function that equals one if doctor i is a member of specialty d, S_d. The model incorporates heterogeneity by allowing different carryover (λ_d) and detailing effectiveness (β_d) across specialties. The estimates for **Equation (15)** using the three estimators are reported in **Table 7**. The general observable pattern with regard to different estimation methods is analogous to the homogeneous model discussed in the previous subsection: presence of serial correlation, biased conventional and restricted estimators, and inefficiency of DGMM due to weak instruments. Thus, for model inference, we again turn our attention to the results from the SGMM estimator based on the restricted-pre method.

[22] We calculated the magnitude of the percentage increase using exp(0.112)-1=0.1185 due to the log-transformed dependent variable.

Two observations are worth noting from **Table 7**: a stronger *long-term* effect (greater inertia) for specialist physicians, and a greater *short-term* marginal effect of detailing for generalist physicians. The parameter estimates of the lagged dependent variable (carryover effect) are larger for specialists whose elasticity measures range from 0.645 to 0.685, than for generalists whose measures range from 0.527 to 0.584; with general surgeons exhibiting the lowest inertia. In contrast to the long-term effect, the parameter estimates associated with the contemporaneous detailing effect are positive and significant (ranging from 0.093 to 0.180) for generalists, whereas those for specialists are small and insignificant. Hence, there is a general trend: specialists exhibit high inertia and low sensitivity to detailing; generalists are less persistent in their prescribing behavior and more responsive to short-term detailing efforts.

6.3. Empirical Evidence: Failure of Tests for Serial Correlation

The results reported in the preceding subsections are conditional on the Arellano-Bond AR(2) tests being rejected (i.e., on second-order serial correlation sufficiently strong to be detected across all methods). Hence, the researcher in this case would have been cautious about applying the conventional dynamic panel data methods and would have utilized restricted moment conditions (lags $t-3$ and earlier) as in our proposed method. This subsection presents a case in which the test statistic built upon conventional methods fail to reject the model despite the presence of serial correlation, leading to biased estimates and incorrect inference.

For this analysis, we run the model in **Equation (14)** separately for each physician specialty. The results for diabetologists appear in **Table 8**. We find that the AR(2) test for the SGMM estimator is rejected only under the restricted or the restricted-pre methods. As is evident from the Monte Carlo experiments, the AR(2) test statistic using conventional moment conditions exhibits weak power and fails to reject the null hypothesis of no second-order serial correlation, as shown in **Table 3a**. The AR(2) test statistic for DGMM under the restricted and the restricted-pre approaches also suffers from both the weak-instruments problem and the moderate effectiveness of explanatory variables, similar to the conditions reported in the far right columns of **Table 3b** and **3c**.

In this case, the conventional test may falsely justify the misspecified model using the unrestricted moment conditions provided by the conventional methods. Thus, by using estimates derived from the conventional methods, researchers can mistakenly infer that sales effort have limited effect (SGMM) or it can even yield a negative outcome (DGMM).

6.4. Discussion

We now return to our main question: How do sales efforts pay off? We address this issue from two perspectives: physician heterogeneity and short- versus long-term trade-off. These perspectives mutually represent significant importance to managerial practice in designing optimal call patterns for route sales.

Our findings reveal substantial variation in detailing effectiveness across different medical practice areas. The results imply that the long-term effect of sales efforts is more pronounced for specialists. Because specialists focus on specific symptoms and prescribe a narrower range of products with only a few substitutes, they commonly exhibit greater stickiness to a particular brand (from a specific firm). In contrast, the short-term marginal effect of detailing is greater for generalists. Because generalists generally prescribe a wide range of generic drugs, many of which have substitutes from competing firms, they are typically more open to prescribing new drugs. Thus, firms should consider the heterogeneity in detailing effectiveness across physician specialties when designing their route sales plans. The appropriate targeting of customers would be especially vital for the generic pharmaceuticals, as in our empirical context, whose success mainly relies on sales volume under tight margins.

In accordance with their sales strategy, firms should set clear objectives for the sales force, for example, between increasing short-term sales versus building long-term relationships. Our results demonstrate a clear trade-off between short- and long-term effectiveness of sales calls. Physicians who tend to exhibit greater persistence are less responsive to detailing, and vice versa. More specifically, specialists exhibit high inertia and low sensitivity to detailing, whereas generalists are more responsive to contemporary detailing efforts and less persistent in their prescribing behavior. The “habit persistence” of physicians’ prescribing behavior stems from costs in learning, searching, and thinking about new scientific information (Janakiraman et al., 2008), which is likely higher for specialist physicians, who possess deeper knowledge in a particular field. However, despite the difficulties in initial influence, these persistent physicians remain attractive in that once captured, they are more likely to remain loyal to the brand they prescribe in the long-run, and competitors' threats would have a limited effect. Thus, these findings serve as a useful supplement for short- and long-term objective setting. By recognizing the trade-off and setting goals that are tailored to each type of physician, firms can increase the overall efficacy of the sales force.

The simple yet robust methodology presented in this paper provides a practical tool for firms in measuring the value of their sales force activity. This method allows researchers (and firms) to control for unobserved heterogeneity, endogeneity issues, and serial correlation, all of which are likely to be a concern in using naturally occurring data. An alternative method to obtain unbiased estimates of sales efforts would be to conduct a controlled field experiment. However, such an experiment would be highly costly for a firm to implement, as it requires random route plan changes that may temporarily forgo firm profit and jeopardize customer relationship. Furthermore, the duration of the experiment must be in considerable length to obtain the long-term outcome of sales efforts, exacerbating the cost issue. Hence, our method of obtaining unbiased estimates of detailing effectiveness from naturally occurring data, in itself, provides significant value to firms, by alleviating the need for costly field studies involving experimentation.

We would like to finalize the discussion by emphasizing the importance of acknowledging the micro-foundation (DGP), which is the fundamental behavioral or theoretical representation of the real-world. The analysis in this study demonstrates that there is value in carefully considering the underlying DGP. It not only provides a robust micro-foundation for causal inference, but also helps unveil the often overlooked underlying assumption or misspecification of the empirical model. As in our case, the micro-foundation of the N-A framework naturally leads to serial correlation in the error structure in applying the dynamic panel data methods.

7. Conclusion

Personal selling in the form of detailing to physicians is the prevailing go-to-market practice in the pharmaceutical industry. Nevertheless, findings on the impact of sales calls have varied widely and controversially, primarily due to inappropriate methods and imprecise data. This paper develops and estimates a generalized model under the Nerlove-Arrow framework to precisely derive the short- and long-term effects of detailing on physicians' prescribing behavior. The dynamic panel data method is utilized to encompass the intertemporal nature of detailing effectiveness, while controlling for physician heterogeneity and correcting for endogeneity issues regarding both lagged dependent and predetermined variables.

We introduce a key methodological insight to the marketing and economics literature. In particular, we challenge the widely used serial correlation assumption (or the lack of such an assumption) about the error structure in applying the conventional dynamic panel data methods, and derive a more appropriate set of moment conditions that can properly address serial correlation. Such correlation is apt to be present in the empirical context of collective marketing efforts over time, characterized by geometric decay. Using the general structure of a dynamic panel data model, this paper reviews the validity of instruments with respect to assumptions about serial correlation and discusses the corresponding plausible moment conditions for estimation.

In addition, we present appropriate moment conditions to properly address the endogeneity concerns arising from the predetermined explanatory variables, an issue that has often been overlooked. The predetermined variable assumption that allows explanatory variables to be correlated with past idiosyncratic shocks is applicable to various settings, especially in marketing, where actions are often a choice of the firm after observing past performance shocks. In the empirical analysis, we further relax this assumption to allow explanatory variables to be weakly predetermined, enabling us to include marketing actions that are simultaneously determined with regard to the current-period idiosyncratic shock.

We also assess the Arellano-Bond specification test for serial correlation, which is the test routinely used in conventional dynamic panel data settings. We provide proof that the test statistic becomes weak and imprecise at detecting serial correlation. This shortcoming leads to a misuse of moment conditions that result in biased parameter estimates and incorrect inference. To validate our claims, we run simulation studies and verify the failure of test statistics under conventional methods. We provide a restricted set of moment conditions that are immune to serial correlation and are appropriate for an unbiased estimation of model primitives.

For the empirical analysis, we apply our proposed method to comprehensive data on sales force detailing. We first show the existence of serial correlation in the data, and the corresponding failure of conventional methods. Inadequate assumptions on serial correlation result in downward bias of parameter estimates. By analyzing differences in the effectiveness of detailing across medical practice areas, we find substantial heterogeneity in both persistence and short-term responsiveness to detailing efforts. Our results reveal that specialist physicians exhibit a greater long-term effect but only modest short-term responsiveness to detailing. In contrast, generalist physicians tend to be more responsive to sales calls in the short term, although the effect may not be long-lasting.

In summary, this paper provides a practical yet rigorous framework to precisely analyze the effectiveness of personal selling efforts. The framework and empirical insights can help firms allocate salesforce resources more efficiently and devise an optimal call-pattern design in route sales. The method can control for various endogeneity concerns that are likely present in naturally occurring data, such as unobserved heterogeneity, state dependence, and serial correlation, without relying on strictly exogenous instruments or controlled field experiments. Although the empirical application presented in this paper is in the personal-selling domain, our model can be extended to other contexts such as advertising. We believe that our proposed method can help both academics and practitioners better understand economic phenomena of a dynamic nature.

References

Acemoglu D, Johnson S, Robinson JA, Yared P (2008) Income and democracy. *American Economic Review*. 98(3): 808-42.

Ahn SC, Schmidt P (1995) Efficient estimation of models for dynamic panel data. *Journal of Econometrics.* 68(1): 5-27.

Amemiya T, MaCurdy TE (1986) Instrumental-variable estimation of an error-components model. *Econometrica*. 54(4): 869-880.

Anderson TW, Hsiao C (1981) Estimation of dynamic models with error components. *Journal of the American Statistical Association*. 76(375): 598-606.

Anderson TW, Hsiao C (1982) Formulation and estimation of dynamic models using panel data. *Journal of Econometrics.* 18(1): 47-82.

Archak N, Ghose A, Ipeirotis PG (2011) Deriving the pricing power of product features by mining consumer reviews. *Management Science*. 57(8): 1485-1509.

Arellano M (2003) *Panel Data Econometrics*. (Oxford University Press).

Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*. 58(2): 277-297.

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

Balestra P, Nerlove M (1966) Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica*. 34(3): 585-612.

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

Bowsher CG (2002) On testing overidentifying restrictions in dynamic panel data models. *Economics Letters*. 77(2): 211-220.

Breusch TS, Mizon GE, Schmidt P (1989) Efficient estimation using panel data. *Econometrica*. 57(3): 695-700.

Burtch G, Ghose A, Wattal S (2013) An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets. *Information Systems Research*. 24(3): 499-519.

Cagan P (1956) The Monetary dynamics of hyperinflation. Friedman M, ed. *Studies in the Quantity Theory of Money* (University of Chicago Press, Chicago), 25-117.

Chung DJ (2017) How much is a win worth? An application to intercollegiate athletics. *Management Science*. 63(2): 548-565.

Clark CR, Doraszelski U, Draganska M (2009) The effect of advertising on brand awareness and perceived quality: An empirical investigation using panel data. *Quantitative Marketing and Economics*. 7(2): 207-236.

Durlauf SN, Johnson PA, Temple JR (2005) Growth econometrics. *Handbook of Economic Growth*. 1:555-677.

Fang E, Lee J, Palmatier R, Han S (2016) If it takes a village to foster innovation, success depends on the neighbors: the effects of global and ego networks on new product launches. *Journal of Marketing Research*. 53(3): 319-337.

Feng H, Morgan NA, Rego LL (2015) Marketing department power and firm performance. *Journal of Marketing*. 79(5): 1-20.

Friedman M (1957) *A Theory of the Consumption Function* (Princeton University Press, Princeton).

Germann F, Ebbes P, Grewal R (2015) The chief marketing officer matters!. *Journal of Marketing*. 79(3): 1-22.

Gönül FF, Carter F, Petrova E, Srinivasan K (2001) Promotion of prescription drugs and its impact on physicians' choice behavior. *Journal of Marketing*. 65(3): 79-90.

Hanataka M (1974) An efficient two-step estimator for the dynamic adjustment model with autoregressive errors. *Journal of Econometrics.* 2(3): 199-220.

Hausman JA, Taylor WE (1981) Panel data and unobservable individual effects. *Econometrica*. 49(6): 1377-1398.

Hirunyawipada T, Xiong G (2018) Corporate environmental commitment and financial performance: Moderating effects of marketing and operations capabilities. *Journal of Business Research*. 86: 22-31.

Hujer R, Rodrigues PJ, Zeiss C (2005) Serial correlation in dynamic panel data models with weakly exogenous regressors and fixed effects. Working Paper, J. W. Goethe-University, Frankfurt.

Janakiraman R, Dutta S, Sismeiro C, Stern P (2008) Physicians' persistence and its implications for their response to promotion of prescription drugs. *Management Science*. 54(6): 1080-1093.

Jung H (2005) A test for autocorrelation in dynamic panel data models. Working Paper, Hitotsubashi University, Tokyo.

Koyck LM (1954) *Distributed Lags and Investment Analysis* (North-Holland Publishing Company, Amsterdam).

Lee N, Moon HR, Zhou Q (2017) Many IVs estimation of dynamic panel regression models with measurement error. *Journal of Econometrics*. 200(2):251-259.

Manchanda P, Chintagunta PK (2004) Responsiveness of physician prescription behavior to salesforce effort: An individual level analysis. *Marketing Letters*. 15(2): 129-145.

Manchanda P, Rossi PE, Chintagunta PK (2004) Response modeling with nonrandom marketing-mix variables. *Journal of Marketing Research*. 41(4): 467-478.

Mathys J, Burmester AB, Clement M (2016) What drives the market popularity of celebrities? A longitudinal analysis of consumer interest in film stars. *International Journal of Research in Marketing*. 33(2): 428-448.

McAlister L, Srinivasan R, Jindal N, Cannella AA (2016) Advertising effectiveness: the moderating effect of firm strategy. *Journal of Marketing Research*. 53(2): 207-224.

Mizik N, Jacobson R (2004) Are physicians "easy marks"? Quantifying the effects of detailing and sampling on new prescriptions. *Management Science*. 50(12): 1704-1715.

Mukherji P, Sorescu A, Prabhu JC, Chandy RK (2011) Behemoths at the gate: How incumbents take on acquisitive entrants (and why some do better than others). *Journal of Marketing*. 75(5): 53-70.

Narasimhan O, Rajiv S, Dutta S (2006) Absorptive capacity in high-technology markets: The competitive advantage of the haves. *Marketing Science*. 25(5): 510-524.

Narayan V, Kadiyali V (2015) Repeated interactions and improved outcomes: an empirical analysis of movie production in the United States. *Management Science*. 62(2): 591-607.

Nerlove M (1958) Distributed lags and demand analysis for agricultural and other commodities. U.S. Department of Agriculture, Washington, DC.

Nerlove M, Arrow KJ (1962) Optimal advertising policy under dynamic conditions. *Economica*. 129-142.

Neumayer E (2004) The impact of political violence on tourism: Dynamic cross-national estimation. *Journal of Conflict Resolution*. 48(2): 259-281.

Parsons LJ, Vanden Abeele, P (1981) Analysis of sales call effectiveness. *Journal of Marketing Research*. 18(1): 107-113.

Paton D (2002) Advertising, quality and sales. *Applied Economics*. 34(4): 431-438.

Rego LL, Morgan NA, Fornell C (2013) Reexamining the market share–customer satisfaction relationship. *Journal of Marketing*. 77(5): 1-20.

Shah D, Kumar V, Kim KH (2014) Managing customer profits: The power of habits. *Journal of Marketing Research*. 51(6): 726-741.

Song R, Meyer J, Ha K (2015) The relationship between product placement and the performance of movies. *Journal of Advertising Research*. 55(3): 322-338.

Staiger D, Stock JH (1997) Instrumental variables regression with weak instruments. *Econometrica*. 65(3): 557-586.

Stock JH, Wright JH, Yogo M (2002) A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics*. 20(4): 518-529.

Terris-Prestholt F, Windmeijer F (2016) How to sell a condom? The impact of demand creation tools on male and female condom sales in resource limited settings. *Journal of Health Economics*. 48: 107-120.

Tuli KR, Bharadwaj SG, Kohli AK (2010) Ties that bind: the impact of multiple types of ties with a customer on sales growth and sales volatility. *Journal of Marketing Research*. 47(1): 36-50.

Van Triest S, Bun MJ, van Raaij EM, Vernooij MJ (2009) The impact of customer-specific marketing expenses on customer retention and customer profitability. *Marketing Letters*. 20(2): 125.

Wall LL, Brown D (2007) The high cost of free lunch. *Obstetrics and Gynecology*. 110(1): 169-173.

Xiong G, Bharadwaj S (2013) Asymmetric roles of advertising and marketing capability in financial returns to news: turning bad into good and good into great. *Journal of Marketing Research*. 50(6): 706-724.

Ye J, Dong B, Lee JY (2017) The long-term impact of service empathy and responsiveness on customer satisfaction and profitability: a longitudinal investigation in a healthcare context. *Marketing Letters*. 28(4): 551-564.

Table 1: Direction of Bias from Invalid Moment Conditions

Invalid Moment Conditions		Direction of Bias			
		λ	β_1	β_2	γ
Lagged Dependent Variables	$\mathrm{E}[y_{i,t-2}\Delta u_{it}] = 0$	Downward	Downward	Downward	No effect
	$\mathrm{E}[u_{it}\Delta y_{i,t-1}] = 0$	Downward	Downward	Downward	Uncertain
Predetermined Variables (Time-varying)	$\mathrm{E}[x_{1it}u_{it}] = 0$	Downward	Downward	Uncertain	Uncertain
	$\mathrm{E}[x_{2i,t-1}\Delta u_{it}] = 0$	Upward	Downward	Downward	No effect
	$\mathrm{E}[u_{it}\Delta x_{2it}] = 0$	Upward	Downward	Downward	Uncertain

Table 2: Types of Estimators and Moment Conditions

Estimator	Moment Conditions		
	Lagged Dependent Variables	Predetermined Variables (Time-varying)	Predetermined Variables (Time-invariant)
Conventional	Equation (7) Equation (8)	Equation (9)	Equation (10)
Restricted	Equation (11) Equation (12)	Equation (9)	Equation (10)
Restricted-pre	Equation (11) Equation (12)	Equation (13)	Equation (10)

Table 3: Simulation Results—Predetermined Explanatory Variable (τ=0.5)

a) Conventional Estimator

Mean Estimates										
True Value	λ	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900
	β	0.900	0.800	0.700	0.600	0.500	0.400	0.300	0.200	0.100
DGMM	λ	0.074	0.143	0.198	0.231	0.232	0.172	0.071	-0.008	-0.023
		(0.013)	(0.017)	(0.019)	(0.024)	(0.029)	(0.037)	(0.038)	(0.035)	(0.030)
	β	0.882	0.772	0.659	0.540	0.414	0.271	0.133	0.032	-0.014
		(0.020)	(0.022)	(0.022)	(0.023)	(0.023)	(0.027)	(0.027)	(0.024)	(0.024)
SGMM	λ	0.073	0.136	0.188	0.223	0.239	0.226	0.185	0.142	0.217
		(0.013)	(0.015)	(0.016)	(0.019)	(0.022)	(0.027)	(0.034)	(0.044)	(0.109)
	β	0.884	0.774	0.662	0.544	0.427	0.304	0.183	0.079	-0.004
		(0.019)	(0.019)	(0.019)	(0.019)	(0.020)	(0.022)	(0.024)	(0.026)	(0.031)
Testing for Serial Correlation (Rejection Frequency in %)										
DGMM	AR(1)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	AR(2)	41.0	77.5	95.0	98.0	100.0	92.5	51.0	4.5	3.5
SGMM	AR(1)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	AR(2)	40.5	76.5	95.0	98.5	100.0	97.5	89.5	63.5	55.0

Note: standard deviation reported in parentheses.

b) Restricted Estimator

Mean Estimates										
True Value	λ	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900
	β	0.900	0.800	0.700	0.600	0.500	0.400	0.300	0.200	0.100
DGMM	λ	0.102	0.211	0.318	0.427	0.532	0.637	0.716	0.471	-0.243
		(0.016)	(0.020)	(0.021)	(0.023)	(0.029)	(0.042)	(0.057)	(0.209)	(0.155)
	β	0.877	0.756	0.637	0.517	0.401	0.285	0.163	0.032	0.035
		(0.020)	(0.021)	(0.020)	(0.019)	(0.019)	(0.020)	(0.020)	(0.024)	(0.045)
SGMM	λ	0.107	0.214	0.324	0.435	0.546	0.663	0.783	0.879	0.966
		(0.015)	(0.018)	(0.019)	(0.022)	(0.029)	(0.037)	(0.054)	(0.072)	(0.076)
	β	0.878	0.758	0.638	0.517	0.401	0.284	0.163	0.036	-0.155
		(0.018)	(0.018)	(0.017)	(0.018)	(0.017)	(0.019)	(0.018)	(0.019)	(0.026)
Testing for Serial Correlation (Rejection Frequency in %)										
DGMM	AR(1)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.0	64.0
	AR(2)	42.0	80.5	96.0	100.0	100.0	100.0	100.0	66.5	29.0
SGMM	AR(1)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	AR(2)	42.5	81.5	97.0	100.0	100.0	100.0	100.0	100.0	99.5

Note: standard deviation reported in parentheses.

c) Restricted-pre Estimator

Mean Estimates										
True Value	λ	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900
	β	0.900	0.800	0.700	0.600	0.500	0.400	0.300	0.200	0.100
DGMM	λ	0.098	0.201	0.301	0.390	0.481	0.560	0.600	0.406	-0.218
		(0.042)	(0.044)	(0.044)	(0.044)	(0.054)	(0.064)	(0.085)	(0.182)	(0.207)
	β	0.893	0.782	0.681	0.602	0.502	0.413	0.322	0.239	0.035
		(0.120)	(0.115)	(0.100)	(0.088)	(0.085)	(0.075)	(0.082)	(0.068)	(0.077)
SGMM	λ	0.100	0.207	0.308	0.400	0.502	0.599	0.702	0.795	0.920
		(0.036)	(0.041)	(0.041)	(0.039)	(0.048)	(0.056)	(0.075)	(0.093)	(0.064)
	β	0.899	0.779	0.676	0.594	0.489	0.392	0.285	0.194	0.085
		(0.109)	(0.108)	(0.093)	(0.083)	(0.075)	(0.071)	(0.075)	(0.063)	(0.066)
Testing for Serial Correlation (Rejection Frequency in %)										
DGMM	AR(1)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.0	23.0
	AR(2)	37.5	76.5	97.0	100.0	100.0	100.0	100.0	73.5	19.0
SGMM	AR(1)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	AR(2)	40.5	79.0	97.0	99.5	100.0	100.0	100.0	100.0	100.0

Note: standard deviation reported in parentheses.

Table 4: Simulation Results—Strictly Exogenous Explanatory Variable (τ=0)

a) Conventional Estimator

Mean Estimates										
True Value	λ	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900
	β	0.900	0.800	0.700	0.600	0.500	0.400	0.300	0.200	0.100
DGMM	λ	0.069	0.132	0.183	0.215	0.220	0.174	0.089	0.013	-0.013
		(0.013)	(0.016)	(0.017)	(0.021)	(0.026)	(0.033)	(0.035)	(0.035)	(0.030)
	β	0.900	0.804	0.702	0.590	0.467	0.320	0.168	0.048	-0.012
		(0.021)	(0.022)	(0.023)	(0.024)	(0.023)	(0.029)	(0.030)	(0.028)	(0.026)
SGMM	λ	0.067	0.124	0.172	0.203	0.220	0.212	0.180	0.143	0.220
		(0.012)	(0.014)	(0.016)	(0.017)	(0.020)	(0.023)	(0.030)	(0.040)	(0.105)
	β	0.901	0.804	0.701	0.590	0.477	0.352	0.226	0.115	0.063
		(0.018)	(0.019)	(0.020)	(0.020)	(0.020)	(0.022)	(0.026)	(0.028)	(0.040)
Testing for Serial Correlation (Rejection Frequency in %)										
DGMM	AR(1)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	AR(2)	26.0	58.0	74.5	84.5	84.0	68.0	27.0	3.5	6.0
SGMM	AR(1)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	AR(2)	27.5	51.5	68.0	82.0	81.5	84.0	69.5	45.5	52.5

Note: standard deviation reported in parentheses.

b) Restricted Estimator

Mean Estimates										
True Value	λ	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900
	β	0.900	0.800	0.700	0.600	0.500	0.400	0.300	0.200	0.100
DGMM	λ	0.096	0.196	0.295	0.392	0.484	0.572	0.634	0.540	-0.146
		(0.015)	(0.018)	(0.018)	(0.023)	(0.029)	(0.041)	(0.053)	(0.139)	(0.214)
	β	0.900	0.800	0.700	0.599	0.499	0.398	0.291	0.159	-0.031
		(0.020)	(0.021)	(0.020)	(0.019)	(0.019)	(0.020)	(0.021)	(0.038)	(0.038)
SGMM	λ	0.100	0.200	0.299	0.400	0.497	0.599	0.702	0.809	0.923
		(0.014)	(0.017)	(0.018)	(0.022)	(0.028)	(0.038)	(0.052)	(0.070)	(0.043)
	β	0.901	0.800	0.699	0.599	0.500	0.401	0.301	0.205	0.104
		(0.018)	(0.018)	(0.017)	(0.017)	(0.016)	(0.018)	(0.017)	(0.018)	(0.019)
Testing for Serial Correlation (Rejection Frequency in %)										
DGMM	AR(1)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	41.0
	AR(2)	43.0	81.5	97.0	100.0	100.0	100.0	100.0	88.0	13.0
SGMM	AR(1)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	AR(2)	43.5	83.0	97.5	100.0	100.0	100.0	100.0	100.0	100.0

Note: standard deviation reported in parentheses.

c) Restricted-pre Estimator

Mean Estimates										
True Value	λ	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900
	β	0.900	0.800	0.700	0.600	0.500	0.400	0.300	0.200	0.100
DGMM	λ	0.081	0.180	0.277	0.366	0.454	0.531	0.565	0.373	-0.204
		(0.039)	(0.042)	(0.046)	(0.045)	(0.055)	(0.070)	(0.095)	(0.186)	(0.219)
	β	0.947	0.844	0.742	0.658	0.554	0.457	0.363	0.264	0.051
		(0.116)	(0.109)	(0.101)	(0.090)	(0.084)	(0.074)	(0.082)	(0.065)	(0.079)
SGMM	λ	0.092	0.194	0.291	0.383	0.481	0.580	0.683	0.781	0.912
		(0.036)	(0.040)	(0.043)	(0.040)	(0.050)	(0.059)	(0.077)	(0.098)	(0.057)
	β	0.924	0.816	0.719	0.635	0.530	0.426	0.319	0.225	0.126
		(0.108)	(0.104)	(0.093)	(0.084)	(0.079)	(0.069)	(0.076)	(0.063)	(0.067)
Testing for Serial Correlation (Rejection Frequency in %)										
DGMM	AR(1)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.5	24.5
	AR(2)	18.5	61.5	85.5	98.5	100.0	99.5	97.5	49.0	18.5
SGMM	AR(1)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	AR(2)	28.5	69.5	91.0	99.0	100.0	100.0	100.0	100.0	100.0

Note: standard deviation reported in parentheses.

Table 5: Descriptive Statistics

a) Physician Prescriptions per Month

	Overall	Cardio-logist	Diabeto-logist	Endocrino-logist	Consulting Physician	General Practitioner	General Surgeon
Mean	19.96	17.64	40.42	32.63	19.87	17.10	17.77
Standard Deviation	32.53	15.08	106.97	40.61	26.80	17.71	14.74
Maximum	1590.00	220.00	1590.00	377.00	1300.00	760.00	108.00
Minimum	1.00	1.00	1.00	1.00	1.00	1.00	1.00
N	9595	628	422	206	4988	3069	282

b) Detailing Calls per Month

	Overall	Cardio-logist	Diabeto-logist	Endocrino-logist	Consulting Physician	General Practitioner	General Surgeon
Mean	2.55	2.70	3.23	4.69	2.68	2.10	2.18
Standard Deviation	1.56	1.67	2.05	2.94	1.55	1.14	1.02
Maximum	24.00	16.00	23.00	24.00	23.00	17.00	7.00
Minimum	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 6: **Estimation Results—Homogeneous Model**

	Conventional		Restricted		Restricted-pre	
	DGMM	SGMM	DGMM	SGMM	DGMM	SGMM
Lagged log (prescription)	**0.214**	**0.262**	**0.459**	**0.599**	**0.485**	**0.562**
	(0.019)	(0.014)	(0.038)	(0.035)	(0.039)	(0.036)
Detailing Calls	**-0.015**	**0.013**	-0.004	**0.017**	0.043	**0.112**
	(0.004)	(0.003)	(0.004)	(0.003)	(0.050)	(0.016)
Specification Tests						
Arellano-Bond AR(1)	Reject	Reject	Reject	Reject	Reject	Reject
Arellano-Bond AR(2)	Reject	Reject	Reject	Reject	Reject	Reject
Number of Instruments	18	28	14	23	13	20
Number of Observations	38,380	47,975	38,380	47,975	38,380	47,975

Dependent variable: logarithm of prescriptions per month. Standard errors are reported in parentheses. Significance (at the 0.05 level) is in bold.

Table 7: Estimation Results—Heterogeneous Model

	Conventional		Restricted		Restricted-pre	
	DGMM	SGMM	DGMM	SGMM	DGMM	SGMM
Lagged log (prescription)						
Cardiologist	0.143	**0.251**	0.260	**0.600**	0.354	**0.685**
	(0.078)	(0.033)	(0.204)	(0.041)	(0.195)	(0.059)
Diabetologist	0.068	**0.394**	0.405	**0.696**	-0.069	**0.645**
	(0.073)	(0.033)	(0.507)	(0.039)	(0.621)	(0.058)
Endocrinologist	0.071	**0.340**	0.293	**0.576**	0.252	**0.652**
	(0.149)	(0.037)	(0.336)	(0.045)	(0.594)	(0.088)
Consulting Physician	**0.200**	**0.270**	**1.148**	**0.603**	**0.744**	**0.584**
	(0.029)	(0.017)	(0.172)	(0.034)	(0.125)	(0.042)
General Practitioner	**0.248**	**0.229**	**-0.300**	**0.546**	0.172	**0.558**
	(0.042)	(0.018)	(0.135)	(0.035)	(0.136)	(0.042)
General Surgeon	0.224	**0.217**	0.140	**0.549**	0.313	**0.527**
	(0.124)	(0.035)	(0.309)	(0.045)	(0.247)	(0.079)
Detailing Calls						
Cardiologist	0.002	0.013	-0.005	0.011	-0.062	-0.001
	(0.015)	(0.011)	(0.017)	(0.012)	(0.077)	(0.043)
Diabetologist	**-0.038**	0.004	-0.017	-0.005	0.048	0.070
	(0.015)	(0.013)	(0.022)	(0.013)	(0.065)	(0.040)
Endocrinologist	0.002	**0.032**	0.010	**0.039**	0.031	0.046
	(0.014)	(0.013)	(0.019)	(0.014)	(0.045)	(0.051)
Consulting Physician	**-0.022**	**0.011**	**0.027**	**0.009**	0.064	**0.093**
	(0.005)	(0.004)	(0.012)	(0.004)	(0.056)	(0.024)
General Practitioner	-0.005	0.007	**-0.040**	**0.017**	-0.048	**0.115**
	(0.008)	(0.006)	(0.012)	(0.006)	(0.087)	(0.036)
General Surgeon	0.022	0.030	-0.007	0.036	**0.146**	**0.180**
	(0.033)	(0.018)	(0.044)	(0.020)	(0.071)	(0.080)
Specification Tests						
Arellano-Bond AR(1)	Reject	Reject	Reject	Reject	Reject	Reject
Arellano-Bond AR(2)	Reject	Reject	Reject	Reject	Reject	Reject
Number of Instruments	88	143	64	113	58	95
Number of Observations	38,380	47,975	38,380	47,975	38,380	47,975

Dependent variable: logarithm of prescriptions per month. Standard errors are reported in parentheses. Significance (at the 0.05 level) is in bold.

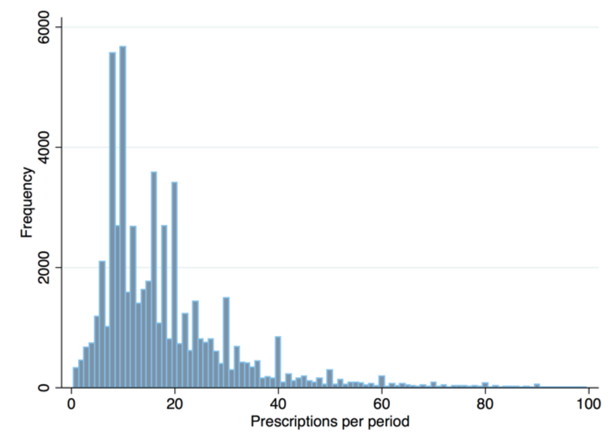
Table 8: Estimation Results—Diabetologists

	Conventional		Restricted		Restricted-pre	
	DGMM	SGMM	DGMM	SGMM	DGMM	SGMM
Lagged log (prescription)	0.119	**0.428**	**0.848**	**0.895**	1.149	**0.773**
	(0.070)	(0.065)	(0.428)	(0.084)	(0.604)	(0.097)
Detailing Calls	**-0.033**	**0.023**	-0.012	0.018	-0.005	**0.072**
	(0.015)	(0.012)	(0.017)	(0.011)	(0.089)	(0.034)
Specification Tests						
Arellano-Bond AR(1)	Reject	Reject	Reject	Reject	Reject	Reject
Arellano-Bond AR(2)	Not Reject	Not Reject	Not Reject	Reject	Not Reject	Reject
Number of Instruments	18	28	14	23	13	20
Number of Observations	1,688	2,110	1,688	2,110	1,688	2,110

Dependent variable: logarithm of prescriptions per month. Standard errors are reported in parentheses. Significance (at the 0.05 level) is in bold.

Figure 1: Distribution of Prescriptions / Calls

a) Number of Prescriptions



b) Number of Calls

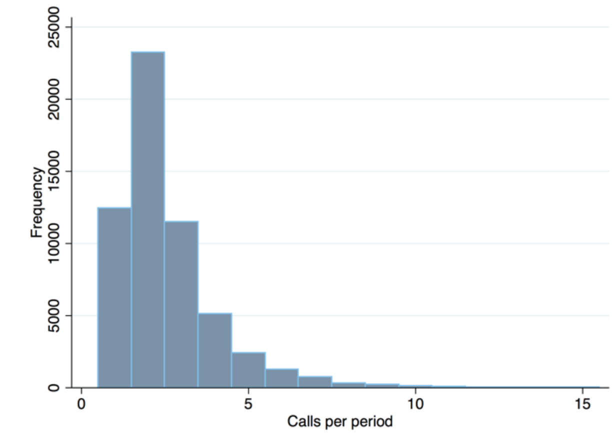
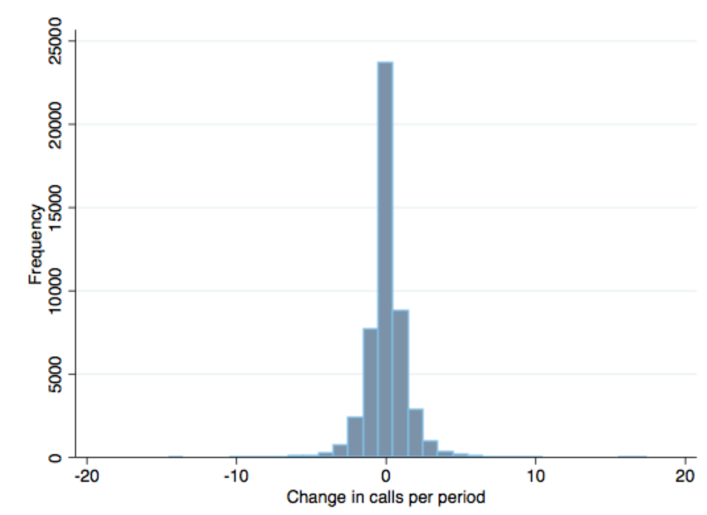


Figure 2: Distribution of Change in Number of Sales Calls



Appendix. Technical Assumptions and Algebraic Derivations

The appendix includes formal statements of assumptions related to the dynamic panel data methods, for both conventional and restricted approaches, and algebraic derivations of the direction of biases from invalid moment conditions.

A. Assumptions Underlying the Dynamic Panel Data Methods

In this section, we provide the assumptions and their implications regarding the dynamic panel data methods discussed throughout the manuscript. We begin with the conventional dynamic panel data methods, followed by restricting the moment conditions under the N-A framework. Subsequently, we incorporate predetermined variables and discuss a generic form of dynamic panel data model under serially correlated errors.

A.1. Conventional Dynamic Panel Data Methods

For expository purposes, we consider a parsimonious dynamic panel data model without explanatory variables of the form:

$$\begin{aligned} y_{it} &= \lambda y_{i,t-1} + u_{it}, \\ u_{it} &= \alpha_i + \varepsilon_{it}. \end{aligned}$$

The model structure has been widely discussed in the economics literature (Anderson and Hsiao, 1981, 1982; Arellano and Bond, 1991; Arellano and Bover, 1995; Ahn and Schmidt, 1995; Blundell and Bond, 1998). The model incorporates the following standard assumptions:

Assumption 1

(A1.1) A random sample of (y_{i1}, y_{i2},..., y_{iT}) for i=1,2,...,N is observed.

(A1.2) $\mathrm{E}[\varepsilon_{it}] = 0$ for all t.

(A1.3) $\mathrm{E}[\varepsilon_{it}\varepsilon_{is}] = 0$ for all $t \neq s$.

(A1.4) $\mathrm{E}[\alpha_i \varepsilon_{it}] = 0$ for all t.

(A1.5) $\mathrm{E}[y_{i1}\varepsilon_{it}] = 0$ for all t.

(A1.6) $|\lambda|<1$.

(A1.7) $\mathrm{E}[\alpha_i \Delta y_{i2}] = 0$.

Here, (A1.1) is the standard i.i.d. assumption, (A1.2) is a common innocuous location normalization, (A1.3) implies that the idiosyncratic shocks are serially uncorrelated, (A1.4) and (A1.5) require the idiosyncratic shocks to be uncorrelated with the individual effect and the initial value of the dependent variable, (A1.6) guarantees that the data are stationary—to rule out the unit-root case, and (A1.7) is the so-called *mean stationarity assumption* concerning the initial condition of the data.

Using assumptions (A1.1)-(A1.6), Arellano and Bond (1991) suggest using the following moment conditions:

(M1.1) $\mathrm{E}[y_{is}\Delta u_{it}] = 0$ for t=3,4,..., T and s=1,2,...,t−2,

which use dependent variables in levels to instrument for the first differenced equation. The conditions provide $\frac{(T-1)(T-2)}{2}$ number of moment conditions.

Ahn and Schmidt (1995) show that when assumption (A1.7) is additionally imposed, the first-period observation of y can be written as

$$y_{i1} = \frac{\alpha_i}{1-\lambda} + \eta_i \, ,$$

where η_i has a zero mean and is uncorrelated with α_i. In other words, the deviation of the first observation from the stationary level $\frac{\alpha_i}{1-\lambda}$ is uncorrelated with the individual effects. Combining the initial condition with the orthogonality to the individual effects assumption (A1.4), we can derive $\mathrm{E}[u_{i3}\Delta y_{i2}] = 0$, and by iteration, the following additional linear moment conditions become further available:

(M1.2) $\mathrm{E}[u_{it}\Delta y_{it-1}] = 0$ for t=3,4,..., T,

which use the lagged differences as instruments for equations in the levels, which provide (T−2) number of additional moment conditions. Blundell and Bond (1998) refer to the GMM estimator based on moment conditions in (M1.1) as the *Difference GMM* (DGMM), and that of the moment conditions in both (M1.1) and (M1.2) as the *System GMM* (SGMM) estimator.

A.2. Restricted Estimator under the Nerlove-Arrow Framework

When the N-A framework is transformed into a dynamic panel data setting, the error structure now becomes

$$y_{it} = \lambda y_{i,t-1} + u_{it},$$
$$u_{it} = \alpha_i + \nu_{it} - \lambda \nu_{i,t-1}.$$

which requires adjusting *assumption 1* to accommodate the new error structure regarding ν_{it} as the following:

Assumption 2

(A2.1) A random sample of (y_{i1}, y_{i2},..., y_{iT}) for i=1,2,...,N is observed.

(A2.2) $\mathrm{E}[\nu_{it}] = 0$ for all t.

(A2.3) $\mathrm{E}[\nu_{it}\nu_{is}] = 0$ for all $t \neq s$.

(A2.4) $\mathrm{E}[\alpha_i \nu_{it}] = 0$ for all t.

(A2.5) $\mathrm{E}[y_{i1}\nu_{it}] = 0$ for all $t \geq 2$.

(A2.6) $|\lambda|<1$.

(A2.7) $\mathrm{E}[\alpha_i \Delta y_{i2}] = 0$.

The rationale behind these assumptions are analogous to *assumption 1*. Note that the mean stationarity assumption of the initial value in (A2.7) is naturally satisfied under the N-A framework, as the model assumes that the data are generated from $t = -\infty$ by construction. Imposing assumptions (A2.1)-(A2.6) leads to the following moment conditions for DGMM under the restricted approach:

(M2.1) $\mathrm{E}[y_{is}\Delta u_{it}] = 0$ for t=4,5,...,T and s=1,2,...,t−3.

These moment conditions correspond to those from (M1.1) with an additional lag. Compared to the moment conditions in (M1.1), the two-period lagged variables become invalid instruments for first-differences. That is, $\mathrm{E}[y_{is}\Delta u_{it}] \neq 0$ when s=t−2—due to serial correlation arising from the N-A framework, which results in $\frac{(T-2)(T-3)}{2}$ number of moment conditions.

When assumption (A2.7) is additionally imposed, the initial condition of the data becomes

$$y_{i1} = \frac{\alpha_i}{1-\lambda} + \nu_{i1} + \eta_i\,,$$

where $\mathrm{E}[\eta_i] = \mathrm{E}[\alpha_i \eta_i] = \mathrm{E}[\nu_{i1}\eta_i] = 0$. However, unlike the conventional approach, $\mathrm{E}[u_{i3}\Delta y_{i2}] \neq 0$ due to the presence of ν_{i2} in both u_{i3} and Δy_{i2}. The lagged differences should also be at least two periods to avoid the common error component. Hence, the previous ($T-2$) number of moment conditions in (M1.2) is replaced by the following ($T-3$) number of conditions for SGMM under the restricted approach:

(M2.2) $\mathrm{E}[u_{it}\Delta y_{it-2}] = 0$ for t=4,5,..., T.

A.3. Incorporating Predetermined Variables

We now return to the model form given by

$$\begin{aligned} y_{it} &= \lambda y_{i,t-1} + \beta' x_{it} + \gamma' z_i + u_{it}, \\ u_{it} &= \alpha_i + \nu_{it} - \lambda \nu_{i,t-1}, \end{aligned}$$

and discuss the assumptions and moment conditions that are pertinent to time-varying explanatory variables x_{it}, and time-invariant explanatory variables z_i. We maintain either *assumption 1* or *assumption 2* for the respective settings. To incorporate the predetermined variables, the following additional assumptions become necessary (Ahn and Schdmidt, 1995; Arellano and Bover, 1995):

Assumption 3

(A3.1) $\mathrm{E}[x_{it}\varepsilon_{is}] = 0$ and $\mathrm{E}[x_{it}\nu_{is}] = 0$ for $t \leq s$.

(A3.2) $\mathrm{E}[z_i\varepsilon_{it}] = 0$ and $\mathrm{E}[z_i\nu_{it}] = 0$ for all t.

(A3.3) $x_{it} = (x_{1it}', x_{2it}')'$ and $\mathrm{E}[x_{1it}\alpha_i] = 0$ for all t.

(A3.4) $z_i = (z_{1i}', z_{2i}')'$ and $E[z_i\alpha_i] = 0$.

Assumptions (A3.1) and (A3.2) imply that the independent variables are predetermined with respect to idiosyncratic shocks. This is weaker than the strict exogeneity assumption, which requires the variables to be uncorrelated with the errors of any time period. Assumptions (A3.3) and (A3.4) are concerned with the correlation between the independent variables and the individual effect.

Given these assumptions, the following moment conditions with regard to the predetermined variables are available for the conventional (and restricted) estimator:

(M3.1) $\mathrm{E}[x_{1it}u_{is}] = 0$ for t=1,2,...,$s-1$ and s=1,2,...,T.

(M3.2) $\mathrm{E}[x_{2it}\Delta u_{is}] = 0$ for t=1,2,...,$s-2$ and s=1,2,...,T.

(M3.3) $\mathrm{E}[z_{1i}u_{it}] = 0$ for t=1,2,...,T.

(M3.4) $\mathrm{E}[z_{2i}\Delta u_{it}] = 0$ for t=2,3,...,T.

For the restricted-pre estimator

(M3.5) $\mathrm{E}[x_{1it}u_{is}] = 0$ for t=1,2,...,s-2 and s=1,2,...,T.

(M3.6) $\mathrm{E}[x_{2it}\Delta u_{is}] = 0$ for t=1,2,...,s-3 and s=1,2,...,T.

(M3.7) $\mathrm{E}[z_{1i}u_{it}] = 0$ for t=1,2,...,T.

(M3.8) $\mathrm{E}[z_{2i}\Delta u_{it}] = 0$ for t=2,3,...,T.

These conditions are derived by exploiting the correlation between the predetermined variable and the error term.

A.4. A Generalized Framework

Here, we consider a more generalized form of the dynamic panel data model, with serially correlated errors in the following general form:

$$
\begin{aligned}
y_{it} &= \lambda y_{i,t-1} + u_{it} \\
u_{it} &= \alpha_i + \nu_{it} + \delta\nu_{i,t-1} + \varepsilon_{it} \, ,
\end{aligned}
$$

where the time-varying idiosyncratic shock at time t is decomposed into two components: (i) ε_{it}, a transitory shock whose effect completely diminishes in the subsequent periods, and (ii) ν_{it}, whose effect persists over to the next period (with decay), imposing a first-order serial correlation. A special case of the above model is when δ=$-\lambda$, which covers applications such as the geometric lag model (N-A framework) in the manuscript or the measurement error model when y_{it} is an observation of y^*_{it} with measurement error ν_{it}, such that $y_{it} = y^*_{it} + \nu_{it}$.

For the above model, *assumptions 1* and *2* are extended to accommodate the new error structure as follows:

Assumption 4

(A4.1) A random sample of (y_{i1}, y_{i2},..., y_{iT}) for i=1,2,...,N is observed.

(A4.2) $\mathrm{E}[\nu_{it}] = \mathrm{E}[\varepsilon_{it}] = 0$ for all t.

(A4.3) $\mathrm{E}[\nu_{it}\nu_{is}] = \mathrm{E}[\varepsilon_{it}\varepsilon_{is}] = 0$ for all $t \neq s$.

(A4.4) $\mathrm{E}[\alpha_i\nu_{it}] = \mathrm{E}[\alpha_i\varepsilon_{it}] = 0$ for all t.

(A4.5) $\mathrm{E}[y_{i1}\nu_{it}] = \mathrm{E}[y_{i1}\varepsilon_{it}] = 0$ for all $t \geq 2$.

(A4.6) $|\lambda|<1$.

(A4.7) $\mathrm{E}[\alpha_i \Delta y_{i2}] = 0$.

(A4.8) $\mathrm{E}[\nu_{it}\varepsilon_{is}] = 0$ for any t and s.

The rationale behind assumptions (A4.1) to (A4.7) are analogous to *assumptions 1* and *2*. Assumption (A4.8) is trivial, which guarantees orthogonality between ν_{it} and ε_{it}. Given assumptions (A4.1)-(A4.8), the moment conditions represented by (M2.1) and (M2.2) can be derived in an analogous manner to section A.2.

B. Direction of Biases from Invalid Moment Conditions

This section provides the algebraic derivations on the direction of biases, for each parameter, arising from the use of invalid moment conditions. For brevity of exposition, we impose the following mild assumptions in determining the signs.

Assumption

(B.1) x_{1it}, x_{2it}, *and* z_i are scalars.

(B.2) $\lambda>0$.

(B.3) $\beta_1>0$ and $\beta_2>0$, which are the coefficients of x_{1it} and x_{2it}, respectively.

(B.4) $\mathrm{E}[x_{1it}\nu_{i,t-1}] > 0$ and $\mathrm{E}[x_{2it}\nu_{i,t-1}] > 0$.

(B.5) $\mathrm{E}[y_{it}y_{i,t-j}]$, $\mathrm{E}[x_{1it}x_{1i,t-j}]$, and $\mathrm{E}[x_{2it}x_{2i,t-j}]$ are positive for any j, and decreases as the lag increases.

The following moment conditions are invalid under the N-A framework:

$$\begin{aligned}
&\mathrm{E}[y_{i,t-2}\Delta u_{it}] = \lambda \mathrm{E}[\nu_{i,t-2}^2] > 0,\\
&\mathrm{E}[u_{it}\Delta y_{i,t-1}] = -\lambda \mathrm{E}[\nu_{i,t-1}^2] < 0,\\
&\mathrm{E}[x_{1it}u_{it}] = -\lambda \mathrm{E}[x_{1it}\nu_{i,t-1}] < 0,\\
&\mathrm{E}[x_{2i,t-1}\Delta u_{it}] = \lambda \mathrm{E}[x_{2i,t-1}\nu_{i,t-2}] > 0,\\
&\mathrm{E}[u_{it}\Delta x_{2it}] = -\lambda \mathrm{E}[x_{2it}\nu_{i,t-1}] < 0.
\end{aligned}$$

By replacing u_{it} in the moment conditions with $y_{it} - \lambda y_{i,t-1} - \beta_1 x_{1it} - \beta_2 x_{2it} - \gamma z_i$, we can rewrite the moment conditions as linear functions of the parameters as follows:

$$\begin{aligned}
&\mathrm{E}[y_{i,t-2}\Delta u_{it}] = \mathrm{E}[y_{i,t-2}\Delta y_{it}] - \lambda \mathrm{E}[y_{i,t-2}\Delta y_{i,t-1}] - \beta_1 \mathrm{E}[y_{i,t-2}\Delta x_{1it}] - \beta_2 \mathrm{E}[y_{i,t-2}\Delta x_{2it}],\\
&\mathrm{E}[u_{it}\Delta y_{i,t-1}] = \mathrm{E}[y_{it}\Delta y_{i,t-1}] - \lambda \mathrm{E}[y_{i,t-1}\Delta y_{i,t-1}] - \beta_1 \mathrm{E}[x_{1it}\Delta y_{i,t-1}] - \beta_2 \mathrm{E}[x_{2it}\Delta y_{i,t-1}] - \gamma \mathrm{E}[z_i\Delta y_{i,t-1}],\\
&\mathrm{E}[x_{1it}u_{it}] = \mathrm{E}[x_{1it}y_{it}] - \lambda \mathrm{E}[x_{1it}y_{i,t-1}] - \beta_1 \mathrm{E}[x_{1it}^2] - \beta_2 \mathrm{E}[x_{1it}x_{2it}] - \gamma \mathrm{E}[x_{1it}z_i],\\
&\mathrm{E}[x_{2i,t-1}\Delta u_{it}] = \mathrm{E}[x_{2i,t-1}\Delta y_{it}] - \lambda \mathrm{E}[x_{2i,t-1}\Delta y_{i,t-1}] - \beta_1 \mathrm{E}[x_{2i,t-1}\Delta x_{1it}] - \beta_2 \mathrm{E}[x_{2i,t-1}\Delta x_{2it}],\\
&\mathrm{E}[u_{it}\Delta x_{2it}] = \mathrm{E}[y_{it}\Delta x_{2it}] - \lambda \mathrm{E}[y_{i,t-1}\Delta x_{2it}] - \beta_1 \mathrm{E}[x_{1it}\Delta x_{2it}] - \beta_2 \mathrm{E}[x_{2it}\Delta x_{2it}] - \gamma \mathrm{E}[z_i\Delta x_{2it}].
\end{aligned}$$

As discussed in section 2.3, we need to check the signs of the derivatives of the moment conditions with respect to the parameters.

B.1. Bias Regarding $\boldsymbol{\lambda}$

The signs of the derivatives of the above linear functions with respect to λ are as follows:

$$\begin{aligned}
&\frac{\partial}{\partial\lambda}\mathrm{E}[y_{i,t-2}\Delta u_{it}] = -\mathrm{E}[y_{i,t-2}\Delta y_{i,t-1}] > 0,\\
&\frac{\partial}{\partial\lambda}\mathrm{E}[u_{it}\Delta y_{i,t-1}] = -\mathrm{E}[y_{i,t-1}\Delta y_{i,t-1}] < 0,\\
&\frac{\partial}{\partial\lambda}\mathrm{E}[x_{1it}u_{it}] = -\mathrm{E}[x_{1it}y_{i,t-1}] < 0,\\
&\frac{\partial}{\partial\lambda}\mathrm{E}[x_{2i,t-1}\Delta u_{it}] = -\mathrm{E}[x_{2i,t-1}\Delta y_{i,t-1}] < 0,\\
&\frac{\partial}{\partial\lambda}\mathrm{E}[u_{it}\Delta x_{2it}] = -\mathrm{E}[y_{i,t-1}\Delta x_{2it}] > 0.
\end{aligned}$$

The first three moment conditions cause a downward bias in λ, but the last two moment conditions lead to an upward bias in λ. Although we cannot determine the sign pre-hoc, our conjecture is that the downward bias would dominate the upward bias. This is because the instrumental variables in y_{it} used in the first two moment conditions are expected to be more relevant than x_{it} used in the latter three moment conditions. Hence, the moment conditions causing the downward bias would receive more weight in the estimation. This is shown in our simulation study in section 4.

B.2. Bias Regarding $\boldsymbol{\beta}$

The signs of derivatives with respect to β_1 are as follows:

$$\begin{aligned}
&\frac{\partial}{\partial \beta_1}\mathrm{E}[y_{i,t-2}\Delta u_{it}] = -\mathrm{E}[y_{i,t-2}\Delta x_{1it}] > 0,\\
&\frac{\partial}{\partial \beta_1}\mathrm{E}[u_{it}\Delta y_{i,t-1}] = -\mathrm{E}[x_{1it}\Delta y_{i,t-1}] < 0,\\
&\frac{\partial}{\partial \beta_1}\mathrm{E}[x_{1it}u_{it}] = -\mathrm{E}[x_{1it}^2] < 0,\\
&\frac{\partial}{\partial \beta_1}\mathrm{E}[x_{2i,t-1}\Delta u_{it}] = -\mathrm{E}[x_{2i,t-1}\Delta x_{1it}] > 0,\\
&\frac{\partial}{\partial \beta_1}\mathrm{E}[u_{it}\Delta x_{2it}] = -\mathrm{E}[x_{1it}\Delta x_{2it}] < 0.
\end{aligned}$$

The direction of the bias in β_1 is coherent, as all invalid moment conditions cause a downward bias.

Similarly, the signs of the derivatives with respect to β_2 are derived as follows:

$$\begin{aligned}
&\frac{\partial}{\partial \beta_2}\mathrm{E}[y_{i,t-2}\Delta u_{it}] = -\mathrm{E}[y_{i,t-2}\Delta x_{2it}] > 0,\\
&\frac{\partial}{\partial \beta_2}\mathrm{E}[u_{it}\Delta y_{i,t-1}] = -\mathrm{E}[x_{2it}\Delta y_{i,t-1}] < 0,\\
&\frac{\partial}{\partial \beta_2}\mathrm{E}[x_{1it}u_{it}] = -\mathrm{E}[x_{1it}x_{2it}],\\
&\frac{\partial}{\partial \beta_2}\mathrm{E}[x_{2i,t-1}\Delta u_{it}] = -\mathrm{E}[x_{2i,t-1}\Delta x_{1it}] > 0,\\
&\frac{\partial}{\partial \beta_2}\mathrm{E}[u_{it}\Delta x_{2it}] = -\mathrm{E}[x_{2it}\Delta x_{2it}] < 0.
\end{aligned}$$

While four of the above five moment conditions generate a downward bias in β_2, the effect of the third moment condition remains uncertain and depends on the relationship between the time-varying explanatory variables.

B.3. Bias Regarding γ

The signs of the derivatives with respect to γ are as follows:

$$\begin{aligned}
&\frac{\partial}{\partial\gamma}\mathrm{E}[y_{i,t-2}\Delta u_{it}] = 0,\\
&\frac{\partial}{\partial\gamma}\mathrm{E}[u_{it}\Delta y_{i,t-1}] = -\mathrm{E}[z_i\Delta y_{i,t-1}],\\
&\frac{\partial}{\partial\gamma}\mathrm{E}[x_{1it}u_{it}] = -\mathrm{E}[x_{1it}z_i],\\
&\frac{\partial}{\partial\gamma}\mathrm{E}[x_{2i,t-1}\Delta u_{it}] = 0,\\
&\frac{\partial}{\partial\gamma}\mathrm{E}[u_{it}\Delta x_{2it}] = -\mathrm{E}[z_i\Delta x_{2it}].
\end{aligned}$$

The effects of the invalid moment conditions on γ are different from those on the other parameters. The first and fourth conditions have no impact on γ, as γ is cancelled out by first-differencing. The other three moment conditions have uncertain signs of the derivative with respect to γ, and depend on their relationship with other explanatory variables. However, if all variables are assumed to be stationary, these three moment conditions would also have no impact on γ.