

Endogeneidade em regressões com dados em painel: Um guia metodológico para pesquisa em finanças corporativas

Lucas A. B. C. Barros¹
Daniel Reed Bergmann¹

¹Universidade de São Paulo, Faculdade de Economia, Administração, Contabilidade e Atuária, São Paulo, Brasil

F. Henrique Castro²

²Fundação Getúlio Vargas, Escola de Economia de São Paulo, São Paulo, Brasil

Alexandre Di Miceli da Silveira³

³Fundação Escola de Comércio Álvares Penteado (FECAP), São Paulo, Brasil

Recebimento:

10/09/2018

Aprovação:

28/06/2019

Editor responsável:

Prof. Dr. Joelson Oliveira
Sampaio

Avaliado pelo sistema:

Double Blind Review

Resumo

Objetivo – Descrever de forma sucinta e prática o uso de defasagens (e/ou diferenças temporais) dos regressores como variáveis instrumentais, mostrando, por meio da discussão teórica ilustrada por um exercício original de simulação, como sua combinação com a adequada modelagem dos efeitos fixos da empresa e do tempo pode endereçar não apenas o problema da endogeneidade dinâmica mas também aqueles decorrentes da presença de variáveis omitidas, erros de mensuração e simultaneidade entre variáveis dependentes e independentes.

Metodologia – Simulação de Monte Carlo.

Resultados – Os tradicionais métodos de MQO, EA e EF podem ser inconsistentes na presença de problemas de endogeneidade bastante plausíveis no contexto de finanças corporativas. Por outro lado, os métodos de estimação para dados em painel baseados no GMM que se utilizam de suposições de exogeneidade sequencial dos regressores apresentam-se como alternativas capazes de contornar eficazmente todos os problemas elencados (desde que as premissas adotadas sejam válidas), mesmo que o pesquisador não disponha de bons instrumentos externos ao modelo.

Contribuições – Discutir e ilustrar via simulações de Monte Carlo um número maior de problemas de endogeneidade, mostrando como eles são endereçados por diferentes estimadores para dados em painel, utilizando uma linguagem menos técnica e mais acessível para pesquisadores ainda não iniciados nos meandros da estimação de modelos dinâmicos para dados em painel.

Palavras-chave – Finanças corporativas; econometria; dados em painel; GMM.



Revista Brasileira de Gestão e Negócios

DOI: 10.7819/rbgn.v22i0.4059

1 Introdução

Grande parte das pesquisas empíricas em finanças corporativas utiliza dados em painel, observando N empresas ao longo de T períodos (tipicamente, com T bem menor do que N). Os dados são oriundos de demonstrações financeiras, cotações de mercado e relatórios da administração, dentre outras fontes, frequentemente com o objetivo de relacionar variáveis e discernir até que ponto uma variável independente (variável explicativa ou regressor) influencia o comportamento da variável dependente (variável resposta). Como exemplo, uma das mais prolíficas linhas de pesquisa nessa tradição é a que procura identificar os determinantes da estrutura de capital das empresas, preocupando-se com as razões pelas quais algumas empresas são relativamente muito endividadas, ao passo que outras utilizam relativamente mais capital próprio para financiar suas atividades (Fama & French, 2002). Outras áreas de investigação analisam os diversos fatores que podem influenciar o valor de mercado, o desempenho financeiro ou o desempenho operacional das empresas. Dentre esses fatores, podem figurar a própria estrutura de capital da empresa, sua estrutura de governança corporativa e características de seus gestores, dentre outras (Bertrand & Schoar, 2003; Himmelberg, Hubbard & Palia, 1999).

Em todos os exemplos acima, o pesquisador está interessado em discernir relações de causalidade entre as variáveis de interesse usando dados reais. Tradicionalmente, a regressão linear é o método aplicado para esse fim. De todos os pressupostos necessários para que uma análise de regressão resulte em inferências adequadas sobre relações de causalidade entre variáveis, o mais importante é o pressuposto da exogeneidade dos regressores. Trata-se do mais difícil de verificar e do mais implausível quando se utilizam dados coletados de empresas. Na prática, essa premissa exclui a possibilidade de correlação diferente de zero entre as variáveis explicativas e o termo de erro do modelo empírico postulado. Se a premissa de não correlação é inválida, diz-se que um ou mais regressores são endógenos. A endogeneidade dos regressores torna inconsistentes os estimadores e

resulta em inferências inadequadas. O problema de endogeneidade no contexto de finanças corporativas decorre normalmente da existência de variáveis omitidas, erros de mensuração das variáveis incluídas no modelo e/ou simultaneidade entre variáveis dependentes e independentes.

A principal vantagem das regressões com dados em painel, que combinam a dimensão transversal e longitudinal, é a possibilidade de modelagem da heterogeneidade não observada (também chamada de efeitos fixos ou efeitos específicos da empresa, supondo que a empresa seja a unidade básica de estudo), representando, por exemplo, características temporalmente estáveis relacionadas à natureza de sua atividade econômica ou à qualidade de sua gestão. Dependendo do contexto da pesquisa, é possível reduzir ou eliminar o problema de endogeneidade decorrente de variáveis omitidas por meio da eliminação da heterogeneidade não observada das unidades observacionais. Há, entretanto, um preço a pagar: em modelos que incluem a heterogeneidade não observada, a consistência do estimador depende da ausência de correlação entre as variáveis explicativas e o termo de erro do modelo em todo e qualquer instante do tempo. Essa condição é conhecida como exogeneidade estrita e é frequentemente ignorada na literatura empírica de finanças corporativas.

O pressuposto de exogeneidade estrita é necessariamente violado quando o modelo inclui defasagens da variável dependente, o que deveria ser bastante comum, como se argumenta neste artigo, considerando a natureza dinâmica da maior parte dos fenômenos de interesse em finanças corporativas (a distorção resultante é conhecida como viés do painel curto, por ser mais acentuada quando T é muito menor do que N , caso típico das pesquisas nessa área). Menos conhecida, todavia, é a violação da exogeneidade estrita decorrente de efeitos de retroalimentação da variável de resposta para os regressores. Esse problema, também conhecido como endogeneidade dinâmica, será frequente em estudos da área, uma vez que choques que afetam a variável dependente (e.g., indicadores de decisões de investimento, de financiamento ou de desempenho financeiro das empresas) possivelmente afetarão quaisquer determinantes dessas variáveis (regressores) em períodos subsequentes.

Uma solução para o problema da endogeneidade dinâmica é o uso de determinadas defasagens (e/ou diferenças temporais) dos regressores originais como variáveis instrumentais, assumindo-se correlação zero entre os instrumentos e os erros do modelo (i.e., premissas de exogeneidade sequencial). O principal objetivo desta pesquisa é descrever de forma sucinta e prática essa estratégia de estimação, mostrando, por meio da discussão teórica ilustrada por um exercício original de simulação, como sua combinação com a adequada modelagem dos efeitos fixos da empresa e do tempo pode endereçar não apenas o problema da endogeneidade dinâmica mas também aqueles decorrentes da presença de variáveis omitidas, erros de mensuração e simultaneidade entre variáveis dependentes e independentes.

Dentre os artigos metodológicos similares a este, com foco em finanças, destacam-se Dang, Kim e Shin (2015), Flannery e Hankins (2013), Wintoki, Linck e Netter (2012) e Zhou, Faff e Alpert (2014). Cada um deles adota um enfoque específico, com diferentes aplicações e enfatizando diferentes aspectos e desafios da estimação de modelos dinâmicos de regressão para dados em painel. Wintoki et al. (2012), por exemplo, enfocam o relacionamento entre a estrutura do conselho de administração e a performance das empresas e não utilizam simulação para comparar o desempenho dos diferentes estimadores em termos de enviesamento e precisão. Flannery e Hankins (2013) e Zhou et al. (2014) utilizam diferentes simulações para comparar o desempenho de estimadores em contextos empíricos similares aos encontrados por pesquisadores da área de finanças corporativas, mas em nenhuma delas modelam a possível determinação simultânea entre a variável de resposta e os regressores. Dang et al. (2015) enfocam a estimação do coeficiente angular da variável dependente defasada e assumem, em suas simulações, que os demais regressores não apresentam problemas de endogeneidade dinâmica ou simultaneidade. Considerando a complexidade da estimação de modelos empíricos com dados observacionais em finanças corporativas, não causa surpresa o fato de que estes artigos chegam em conclusões e recomendações por vezes diferentes, sem que

se possa identificar uma estratégia de estimação uniformemente superior.

O presente estudo se diferencia dos demais, em primeiro lugar, por discutir e ilustrar via simulações de Monte Carlo, em conjunto e separadamente, um número maior de problemas de endogeneidade (i.e., efeitos de retroalimentação, variáveis omitidas, erros de mensuração e simultaneidade), mostrando como eles são endereçados por diferentes estimadores para dados em painel. Em particular, este é o único estudo, até onde sabemos, a modelar explicitamente o chamado efeito fixo do tempo, mostrando que sua omissão pode introduzir um relevante viés de variável omitida. Em segundo lugar, este artigo utiliza uma linguagem menos técnica e mais acessível para pesquisadores ainda não iniciados nos meandros da estimação de modelos dinâmicos para dados em painel. Por outro lado, esta pesquisa é mais simples tecnicamente do que as supracitadas e não discute em detalhe as dificuldades técnicas da aplicação dos estimadores avaliados quando os pressupostos que garantem sua correção são violados, por exemplo, em razão da censura da variável dependente ou da presença de autocorrelação nos erros do modelo. Portanto, este trabalho pode servir como uma referência complementar para os pesquisadores, mas sem pretensão de substituir outros guias metodológicos.

A discussão teórica sugere que problemas de endogeneidade devem afligir muitos estudos empíricos com dados observacionais em finanças corporativas e as simulações mostram que tais problemas podem prejudicar substancialmente as inferências baseadas em estimadores que não são capazes de endereçá-los adequadamente. Em particular, este estudo alerta para a possível inconsistência, em muitos contextos de interesse, dos tradicionais estimadores de Mínimos Quadrados Ordinário (MQO), Efeitos Aleatórios (EA) e Efeitos Fixos (EF). Por outro lado, certos estimadores para painel baseados no Método dos Momentos Generalizado (GMM), a exemplo daquele conhecido como GMM Sistêmico (ou Blundell-Bond), são capazes, em modelos cuidadosamente especificados, de endereçar os principais problemas de endogeneidade e dessa

forma produzir inferências mais adequadas mesmo na ausência de experimentos naturais ou de variáveis instrumentais externas ao modelo. Não obstante, a consistência de qualquer estimador depende da validade dos pressupostos a ele subjacente. Embora os pressupostos dos estimadores GMM referidos acima sejam frequentemente menos restritivos e mais plausíveis do que os de estimadores mais tradicionais, a literatura econométrica mostra que sua violação pode distorcer substancialmente as inferências (Dang et al., 2015; Wintoki et al., 2012). Além disso, limitações dos dados e problemas na especificação das regressões podem ocasionar substancial enviesamento em amostras finitas (i.e., amostras relativamente pequenas. Vide, por exemplo, Windmeijer, 2018; Bun & Sarafidis, 2015).

O artigo é estruturado da seguinte forma: a seção 2 discute as principais causas do problema de endogeneidade no contexto de finanças corporativas e o uso de variáveis instrumentais como solução genérica para esse problema; a seção 3 discute os métodos de regressão para dados em painel mais comumente usados na pesquisa empírica em finanças corporativas e empregados nas simulações do presente artigo; a seção 4 apresenta e discute os resultados do exercício de simulação; e a seção 5 conclui o trabalho.

2 Fontes de endogeneidade e variáveis instrumentais

Considere o modelo linear a seguir:

$$y_i = \alpha + \beta x_i + \varepsilon_i, i = 1, \dots, N, \quad [1]$$

tal que i corresponde à i -ésima empresa de uma amostra aleatória contendo N empresas, y é a variável resposta, x é o regressor de interesse e ε é o termo de erro aleatório. Suponha que o parâmetro β represente o efeito causal (linear, neste exemplo) de x (e.g., porte da empresa, suas práticas de governança corporativa, alavancagem etc.) sobre y (e.g., desempenho financeiro da empresa, sua estrutura do conselho de administração etc.).

Para que β seja estimado de forma consistente (i.e., com viés tendendo a zero na medida em que N aumenta), um pressuposto fundamental é o da não correlação entre x e ε , caso em que x seria definido como um regressor exógeno. A suposição de exogeneidade, entretanto, não pode ser facilmente verificada, uma vez que, diferentemente de x e de y , ε não é diretamente observável. As causas fundamentais que levam à violação desse pressuposto são bem conhecidas e discutidas a seguir.

2.1 Fontes de endogeneidade

2.1.1 Variáveis omitidas

Talvez a causa mais comum (ou mais evidente) de endogeneidade em modelos de regressão seja a omissão de variáveis simultaneamente correlacionadas com os regressores incluídos e com a variável resposta. Na equação (1), o problema pode ser representado por uma variável ω que influencia o comportamento de y e de x ao mesmo tempo. Sua omissão em (1) significa que ω será incorporada ao erro ε , induzindo alguma correlação entre ε e a variável de interesse x e tornando inconsistente a estimação de β . Uma solução padrão para o problema seria incluir ω entre os regressores, aumentando dessa forma o modelo original, como mostrado abaixo:

$$y_i = \alpha + \beta x_i + \delta \omega_i + \varepsilon_i. \quad [2]$$

Nesse caso, ω seria considerada uma variável de controle. A inclusão de variáveis de controle (e.g., $\omega_1, \dots, \omega_k$) nas regressões tem sido a forma preferencial para evitar possíveis problemas de endogeneidade nos estudos empíricos de finanças corporativas (ω também pode ser, naturalmente, uma transformação de x , a exemplo de x^2 ou x^3 , visando capturar relacionamentos não lineares entre x e y , por exemplo). Essa estratégia não funcionará se ω for intrinsecamente não mensurável ou se o pesquisador não dispõe de informações suficientes para mensurá-la de maneira fidedigna. Infelizmente, pode-se esperar que isso ocorra em

boa parte (senão na maior parte) dos estudos empíricos dessa área de pesquisa.

Não é difícil pensar em exemplos de variáveis omitidas não observáveis (ou não mensuradas) no contexto de finanças corporativas. Nesse caso, ω poderia representar a habilidade dos gestores, elementos da cultura organizacional ou vantagens competitivas da empresa possivelmente correlacionadas com x e y . Mesmo variáveis potencialmente mensuráveis como o poder de mercado da empresa, o qual poderia influenciar simultaneamente seu desempenho financeiro, valor de mercado, estrutura de financiamentos, oportunidades de crescimento e práticas de governança corporativa, dentre outros indicadores de interesse para finanças, frequentemente são ignoradas em estudos empíricos em razão da indisponibilidade de dados ou da dificuldade de construção de variáveis substitutas (*proxies*) que efetivamente capturem o fenômeno.

2.1.2 Erros de mensuração

Em estudos com dados observacionais de empresas, é razoável supor que tanto y quanto x possam ser mensurados com algum grau de imprecisão, causada tanto por erros de registro (e.g., digitação equivocada ou arredondamentos) como pelo distanciamento entre um construto que se desejaria observar e a *proxy* efetivamente disponível. Genericamente, podemos representar o problema por meio da equação:

$$x_i = x_i^* + e_i, i = 1, \dots, N, \quad [3]$$

tal que x_i é a variável efetivamente observada, x_i^* seu valor “verdadeiro” e e_i o erro de mensuração, ou ruído. Um raciocínio análogo se aplica à variável y .

Normalmente, os argumentos teóricos que orientam a formulação dos modelos empíricos postulam determinados relacionamentos entre construtos (e.g., valor, desempenho, porte, qualidade das práticas de governança corporativa etc.), os quais frequentemente não correspondem de forma exata aos indicadores observados pelo

pesquisador. Em outras palavras, suponha que o modelo que se gostaria de estimar seja:

$$y_i^* = \alpha + \beta x_i^* + \varepsilon_i, \quad [4]$$

mas que apenas as medidas y_i e x_i , possivelmente mensuradas com erro, estejam disponíveis. Essa é uma dificuldade certamente comum a muitos estudos empíricos no campo das finanças corporativas, e seus efeitos sobre as estimativas resultantes dependem de suposições sobre o comportamento dos erros de mensuração.

Suponha que apenas x é mensurado com erro e que o modelo que se gostaria de estimar é $y_i = \alpha + \beta x_i^* + \varepsilon_i$. Uma vez que x_i^* não é observável, a equação efetivamente estimada, substituindo na equação acima a equação (3), será:

$$\begin{aligned} y_i &= \alpha + \beta x_i^* + \varepsilon_i \\ &= \alpha + \beta(x_i - e_i) + \varepsilon_i \\ &= \alpha + \beta x_i + u_i, \end{aligned} \quad [5]$$

tal que $u_i \equiv \varepsilon_i - \beta e_i$ é o termo de erro do modelo efetivamente estimado. Nesse caso, β será estimado de forma consistente se u e x forem não correlacionados. Para tanto, é necessário que não exista correlação entre e e x , e também entre o erro de mensuração e e x . Infelizmente, ainda que a primeira suposição seja válida, em muitos casos a segunda não será. Como ilustração, x pode ser o valor de mercado observado da empresa, x^* a parcela de x determinada pelos fundamentos do negócio avaliados pelos investidores e e a parcela do preço devida a diversas formas de ruído, incluindo movimentos especulativos. Os erros de apreçamento agregados em podem ser independentes dos fundamentos da empresa, mas provavelmente correlacionam-se positivamente com o valor de mercado observado pelo pesquisador.

Quando e e x são correlacionados, os estimadores tradicionais para os parâmetros da equação (5) tornam-se inconsistentes. Mais especificamente, é provável que o valor estimado

para o coeficiente β seja inferior, em módulo, a seu valor verdadeiro (aquele que seria obtido se x fosse mensurado sem erro), fenômeno conhecido como viés de atenuação. Se vários regressores, entretanto, contiverem erros de mensuração correlacionados com seus valores observados, a direção da inconsistência resultante costuma ser indeterminada (Greene, 2000). Um raciocínio análogo se aplica ao erro de mensuração em y correlacionado com x . Em ambos os casos, a inconsistência resultante se assemelha à causada por variáveis omitidas (para uma discussão detalhada, vide Roberts & Whited, 2013).

2.1.3 Simultaneidade

Uma fonte comum de problemas de endogeneidade em pesquisas na área de finanças corporativas é a provável determinação simultânea de diversas variáveis. De fato, considerando-se a complexa interdependência das decisões corporativas, pode-se argumentar que esta deveria ser uma preocupação de primeira ordem para pesquisadores empíricos da área. Um exemplo é o relacionamento entre a alavancagem e o valor de mercado das empresas. Diferentes argumentos teóricos levam a crer que medidas de valor de mercado, como *proxies* para as oportunidades futuras de investimento disponíveis, podem influenciar contemporaneamente a política de financiamento das empresas (Fama & French, 2002). Ao mesmo tempo, outras linhas de argumentação sugerem que a alavancagem pode exercer influência sobre o desempenho da organização, por exemplo, por meio da redução de seu caixa disponível, o qual poderia ser utilizado ineficientemente por gestores autointeressados, contribuindo, em parte, para a determinação do valor de mercado da empresa (McConnell & Servaes, 1995; Stulz, 1990). Raciocínios análogos podem ser aplicados a muitas outras variáveis corporativas, tornando ambíguo o sentido das relações de causalidade esperadas.

Uma eventual simultaneidade (também conhecida como determinação simultânea ou causalidade reversa) na relação entre y e x , de tal forma que ambas as variáveis podem ser consideradas independentes ou dependentes uma

em relação à outra, introduzirá correlação entre o regressor e o erro do modelo, novamente tornando viesados e inconsistentes os estimadores de β que ignoram o problema.

2.2 Variáveis instrumentais e quase-experimentos

A solução genérica para qualquer problema de endogeneidade, seja ele causado por erros de mensuração, variáveis omitidas ou simultaneidade, é a utilização de variáveis instrumentais válidas. Voltando ao modelo inicial $y_i = \alpha + \beta x_i + \varepsilon_i$, a variável x será endógena se for correlacionada com ε . Esse problema impossibilitará a estimação consistente do parâmetro de interesse β , a menos que exista outra variável z que seja, ao mesmo tempo, correlacionada com x e não correlacionada com ε . Portanto, com respeito ao modelo acima, z seria uma variável exógena. Nesse caso, uma possibilidade é implementar uma estimação em dois estágios ilustrada a seguir. Primeiro, estimam-se os parâmetros do modelo que relaciona x e z :

$$x_i = \delta_1 + \delta_2 z_i + v_i, i = 1, \dots, N, \quad [6]$$

assumindo-se que $\delta_2 \neq 0$. Em seguida, os parâmetros estimados ($\hat{\delta}_1$ e $\hat{\delta}_2$) são utilizados para construir uma variável (\hat{x}_i) resultante da projeção de x em z , tal que $\hat{x}_i \equiv \hat{\delta}_1 + \hat{\delta}_2 z_i$. Logo, \hat{x}_i corresponde aos valores ajustados ou previstos por essa primeira regressão linear.

Na segunda etapa, substitui-se a variável original x por \hat{x} e estima-se a equação (7) a seguir:

$$y_i = \alpha + \beta \hat{x}_i + \varepsilon_i, i = 1, \dots, N. \quad [7]$$

Uma vez que não há correlação entre \hat{x} e ε , também não existirá correlação entre z e ε . De fato, \hat{x} pode ser entendido como a parcela de x não correlacionada com ε . Quando mais de um instrumento exógeno para x está disponível, a extensão do procedimento descrito acima é natural, bastando incluí-los como regressores na equação (6). Não obstante a simplicidade dessa estratégia de identificação, o grande desafio para

os pesquisadores é encontrar um instrumento ou conjunto de instrumentos válidos e suficientemente correlacionados com as variáveis endógenas. Essa dificuldade se agrava porque, embora a primeira suposição, de correlação significativa entre os instrumentos e o regressor endógeno, seja verificável, a segunda, de não correlação deles com o termo de erro do modelo, não o é, uma vez que o erro não é diretamente observável. Larcker e Rusticus (2010) discutem, no contexto da pesquisa contábil, similar ao da pesquisa em finanças corporativas, os principais problemas e desafios das estratégias de identificação que utilizam variáveis instrumentais externas ao modelo, destacando o problema dos instrumentos fracos e a provável endogeneidade de muitos instrumentos propostos na literatura.

Com o intuito de aumentar a credibilidade de suas estratégias de identificação, uma parcela crescente da literatura de finanças corporativas utiliza variáveis instrumentais derivadas de contextos peculiares, genericamente denominados experimentos naturais ou quase-experimentos. Em sua maioria, essas pesquisas exploram peculiaridades ou eventos aparentemente exógenos (portanto, não influenciados pelas próprias variáveis corporativas de interesse), incluindo mudanças em leis e regulamentos impostos a um conjunto de empresas. Além da construção de variáveis instrumentais, contextos quase-experimentais ensejam o uso de outras estratégias de identificação de efeitos causais, com destaque para estudos de evento, regressões com descontinuidade, modelos de diferença em diferenças e *propensity score matching* (vide, por exemplo, Roberts & Whited, 2013; Angrist & Pischke, 2008). Em alguns casos, duas ou mais dessas estratégias são utilizadas simultaneamente com o intuito de mitigar problemas de endogeneidade. Black e Kim (2012), por exemplo, explorando uma mudança regulatória na Coreia do Sul e o fato dessa mudança se aplicar exclusivamente a empresas de grande porte empregam variáveis instrumentais, estudo de evento e análises de descontinuidade e de diferença em diferenças para estudar a influência da estrutura do conselho de administração sobre o valor de mercado das empresas.

Pesquisas em finanças corporativas que empregam variáveis instrumentais e/ou quase-experimentos frequentemente utilizam também métodos de estimação para dados em painel como parte de sua estratégia empírica (Black & Kim, 2012). Não obstante, os procedimentos de estimação discutidos nas próximas seções têm como principal atrativo a possibilidade de mitigação de problemas de endogeneidade dos regressores mesmo na ausência de instrumentos externos ao modelo e de contextos quase-experimentais, ausência esta comum a grande parte dos estudos empíricos em finanças corporativas.

3 Métodos de regressão para dados em painel

Acrescentando à equação geral (1) a dimensão longitudinal, representamos na forma abaixo o modelo empírico de interesse:

$$y_{it} = \alpha + \beta x_{it} + \varepsilon_{it}, i = 1, \dots, N, t = 1, \dots, T. [8]$$

A única diferença entre (1) e (8) é que, agora, as N empresas são observadas ao longo de T períodos, de tal forma que os subscritos i e t representam, respectivamente, a i -ésima empresa e o t -ésimo momento do tempo. A seguir, discutimos as possibilidades de modelagem oferecidas por painéis e seus benefícios potenciais no controle dos problemas de endogeneidade. De forma geral, os procedimentos apresentados abaixo são apropriados para painéis curtos, entendidos como aqueles nos quais N é bem maior que T , como é o caso da maior parte das amostras disponíveis para os pesquisadores de finanças corporativas. Assim, todos os resultados assintóticos aplicáveis à discussão abaixo baseiam-se na suposição de que T é fixo e $N \rightarrow \infty$ (ou, menos formalmente, é fixo T e N é “grande”).

3.1 A heterogeneidade não observada

Uma das mais interessantes possibilidades oferecidas por amostras dispostas em painel é a modelagem explícita de variáveis que não são observadas pelo pesquisador (seja por falta de informações, seja porque tais variáveis são

intrinsecamente não observáveis). Esse novo componente pode ser representado como uma decomposição do termo de erro da equação (8),

$$y_{it} = \alpha + \beta x_{it} + \eta_i + u_{it}, i = 1, \dots, N, t = 1, \dots, T, \quad [9]$$

tal que η_i representa a heterogeneidade não observada das empresas da amostra e u_{it} o termo de erro do modelo. A única restrição sobre o comportamento de η_i é que ele deve variar apenas entre empresas e não ao longo do tempo. Na prática, isso significa que η_i captura toda e qualquer heterogeneidade não observada associada à empresa i que seja invariante ao longo do período amostral. No contexto de finanças corporativas, isso pode incluir elementos da cultura organizacional da empresa, a habilidade ou capital intelectual de seus colaboradores, sua capacidade de inovação, além de outras vantagens competitivas e idiosincrasias, inclusive vinculadas à natureza de sua atividade empresarial, na medida em que estas sejam estáveis no tempo ou, pelo menos, ao longo do período amostral.

Dependendo do método utilizado para estimar os parâmetros do modelo (9), a inclusão de η_i pode ajudar a reduzir ou eliminar o problema de variáveis omitidas, tão comum em muitos contextos empíricos de interesse para finanças corporativas, complementando de maneira eficaz a tradicional inclusão de variáveis de controle (nesse caso, apenas variáveis de controle que variam ao longo do período amostral precisariam ser incluídas).

A estimação de modelos contendo η_i pode ser conduzida de formas diversas, dependendo dos objetivos da pesquisa e das suposições adotadas pelo pesquisador. Os diferentes procedimentos são frequentemente agrupados em duas categorias: Efeitos Aleatórios (EA) ou Efeitos Fixos (EF)¹. Em ambos os casos, a estimação consistente de β depende fundamentalmente da premissa de não correlação entre o erro u_{it} e o regressor de interesse x , observado em qualquer instante de tempo. Logo, assume-se não apenas a não correlação entre u_{it} e x_{it} , mas entre u_{it} e x_{i1}, \dots, x_{iT} . A abordagem de EA, todavia, impõe a suposição

na forma $\varepsilon_{it} = \eta_i + u_{it}$, resultando no modelo estendido abaixo:

adicional de não correlação entre x_{i1}, \dots, x_{iT} e o efeito específico o η_i . Em termos da identificação do parâmetro β , esta pode ser considerada a diferença fundamental entre as duas abordagens². Se a premissa de não correlação entre η e x for considerada irrealista, os procedimentos de EF serão, em princípio, mais adequados.

3.2 A suposição de exogeneidade estrita e efeitos de *feedback*

O pressuposto fundamental para a correta estimação dos parâmetros de modelos com heterogeneidade não observada utilizando os procedimentos tradicionais de EF ou EA pode ser mais restritivo do que parece e merece um exame específico. Para facilitar a exposição, afirmações acerca da correlação entre erros e regressores serão substituídas por afirmações acerca da esperança condicional dos erros. Assim, o pressuposto fundamental para a estimação dos parâmetros da equação (8) através dos procedimentos de EF ou EA pode ser formalizado como:

$$E(u_{it} | x_{i1}, x_{i2}, \dots, x_{iT}, \eta_i) = 0, \quad [10]$$

tal que $E(\cdot)$ é o operador de valor esperado. A expressão acima é conhecida como a suposição de exogeneidade estrita dos regressores e é uma condição suficiente para a não correlação entre u_{it} e x_{i1}, \dots, x_{iT} . O pressuposto de exogeneidade estrita dos regressores descarta qualquer possibilidade de correlação entre os erros contemporâneos e valores passados, contemporâneos ou futuros das variáveis explicativas. Embora esta seja uma premissa aceitável em alguns contextos de pesquisa, em outros ela será irrealista.

Considere, como ilustração, um típico modelo de finanças corporativas com o grau de alavancagem da empresa sendo explicado por

sua lucratividade e por seu valor de mercado. O termo de erro dessa regressão capturar todos os choques que podem afetar contemporaneamente a alavancagem, por exemplo, uma mudança na condução estratégica do negócio que implique, dentre outras coisas, a reorganização imediata de sua estrutura de financiamentos. Ainda que essa mudança estratégica não influencie contemporaneamente os regressores, isto é, a lucratividade e o valor de mercado, é bastante provável que ela se correlacione com seus valores futuros. Esse fenômeno é conhecido como a retroalimentação da variável de resposta para os regressores (efeito *feedback*) no sentido de que, voltando ao exemplo, alterações da alavancagem podem influenciar a lucratividade e o valor de mercado futuros da organização. Se houver retroalimentação, a suposição de exogeneidade estrita não será atendida, tornando inconsistentes os estimadores tradicionais de EF e EA.

De fato, em face da interdependência das decisões corporativas, é prudente esperar algum grau de retroalimentação da variável dependente para os regressores em quase todos os contextos empíricos de interesse para pesquisadores de finanças corporativas. Esse fenômeno, geralmente ignorado em estudos empíricos da área que se utilizam de dados em painel, é bem discutido por Wintoki et al. (2012). Nele, os autores se referem ao problema como endogeneidade dinâmica e oferecem exemplos de sua ocorrência no contexto de pesquisas que relacionam desempenho empresarial e governança corporativa.

O problema descrito acima pode ser resolvido utilizando-se quaisquer estimadores de EF ou EA adaptados para acomodar variáveis instrumentais, desde que bons instrumentos estritamente exógenos estejam disponíveis. Alternativamente, alguns procedimentos, como se discute a seguir, permitem a estimação consistente de modelos com heterogeneidade não-observada utilizando instrumentos baseados em defasagens dos regressores originais e premissas bem menos restritivas do que a formalizada em (10).

É importante observar que formulações que ignoram a heterogeneidade não observada, do tipo $y_{it} = \alpha + \beta x_{it} + \varepsilon_{it}$, cujos parâmetros são tipicamente estimados por MQO aplicado

a dados em painel (também conhecido como MQO para dados agrupados, ou *Pooled OLS*), utilizam como pressuposto fundamental o da não correlação contemporânea entre os erros e os regressores, cuja condição suficiente é representada pela equação (11):

$$E(\varepsilon_{it} | x_{it}) = 0. \quad [11]$$

Essa premissa é bem menos restritiva do que a de exogeneidade estrita. Ou seja, nesse contexto a presença de efeitos de *feedback* não tornará o regressor endógeno. Por outro lado, é claro, o pressuposto em (11) será violado se em ε_{it} houver um efeito específico correlacionado com os regressores.

3.3 Procedimentos baseados no método dos momentos generalizado

A discussão acima sugere que os modelos empíricos que consideram explicitamente a heterogeneidade não observada das empresas são provavelmente mais adequados para descrever o relacionamento estatístico entre as variáveis de interesse em pesquisa no campo de finanças corporativas. Os métodos mais comumente empregados para estimar modelos desse tipo, normalmente classificados como estimadores de EA ou de EF, exigem, entretanto, que os regressores sejam estritamente exógenos, premissa provavelmente muito restritiva em estudos que utilizam dados de empresas e que será violada se houver retroalimentação da variável de resposta para os regressores. Naturalmente, as outras fontes potenciais de problemas de endogeneidade, apresentadas nas seções anteriores, também podem contribuir para a violação desse pressuposto.

Uma solução natural para esse problema é o uso de variáveis instrumentais externas ao modelo de interesse. É teoricamente possível, por exemplo, encontrar instrumentos estritamente exógenos para cada um dos regressores suspeitos de endogeneidade. Na prática, todavia, variáveis com essas características e que ainda apresentem forte correlação com os regressores normalmente não estão disponíveis em estudos de finanças

corporativas, pelo menos em número suficiente. Os métodos descritos nesta seção, por outro lado, permitem o uso de instrumentos apenas sequencialmente exógenos, baseados, por exemplo (mas não necessariamente), em defasagens adequadas dos próprios regressores originais.

Considere novamente a formulação mostrada em (9). Suponha que x seja correlacionado (por efeito de retroalimentação) com os valores passados dos termos de erro ($u_{it-1}, u_{it-2}, \dots, u_{i1}$), mas que ele não se correlacione com seus valores contemporâneos ou futuros. Uma condição suficiente para essa última suposição pode ser expressa na forma:

$$E(u_{it}|x_{i1}, x_{i2}, \dots, x_{it}, \eta_i) = 0. \quad [12]$$

Diz-se, nesse caso, que x é sequencialmente exógeno, por oposição ao pressuposto mais restritivo de exogeneidade estrita formalizado pela equação (10) (Wooldridge, 2010). A ideia de exogeneidade sequencial pode ser naturalmente estendida para acomodar quaisquer defasagens ou valores futuros dos regressores supostamente não correlacionados com os erros. A determinação simultânea dos regressores e da variável de resposta, por exemplo, pode induzir alguma correlação entre x_{it} e u_{it} . Nesse caso, o pressuposto (12) não será válido, mas a suposição

$$E(u_{it}|x_{i1}, x_{i2}, \dots, x_{it-1}, \eta_i) = 0, \quad [13]$$

será adequada se não houver correlação entre os regressores e os valores futuros do termo de erro do modelo. Problemas similares de endogeneidade podem ser causados pela presença de erros de mensuração em x_{it} e sua solução também pode

passar por suposições de exogeneidade sequencial dos regressores³.

Diversos métodos de estimação apropriados para painéis curtos e que utilizam variáveis sequencialmente exógenas como instrumentos estão disponíveis e são por vezes classificados em dois grupos: estimadores de Variáveis Instrumentais e estimadores baseados no Método dos Momentos Generalizado (GMM). Esses métodos foram desenvolvidos, *grosso modo*, tendo como foco a estimação de modelos dinâmicos. Entenda-se por especificações dinâmicas, nesse contexto, modelos empíricos que incluem entre os regressores uma ou mais defasagens da variável de resposta, tipicamente apenas a primeira defasagem. Em outras palavras, numa formulação como a mostrada em (9), incluir-se-ia y_{it-1} entre os regressores e, por definição, y_{it-1} não é uma variável estritamente exógena. Não obstante, os métodos aqui discutidos são igualmente válidos para modelos estáticos como o mostrado em (9), ou seja, formulações que não incluem defasagens de y_{it} entre os regressores. Uma boa introdução a essa literatura é oferecida por Bond (2002).

Dentre os diversos métodos desenvolvidos para painel capazes de incorporar variáveis instrumentais dois se destacam, em razão de sua eficiência e flexibilidade para acomodar diferentes padrões de comportamento das variáveis de interesse. O primeiro é um procedimento desenvolvido por Arellano e Bond (1991) e apelidado de estimador Arellano-Bond ou GMM em Diferenças (GMM-Dif).

Esse procedimento primeiro transforma as variáveis do modelo com o intuito de eliminar a heterogeneidade não observada. Para tanto, a transformação normalmente aplicada consiste em calcular as diferenças das variáveis com relação a seus valores defasados. Aplicando essa transformação ao modelo (9), obtém-se a equação (14):

$$\Delta y_{it} = \beta \Delta x_{it} + \Delta u_{it}, i = 1, \dots, N, t = 1, \dots, T, \quad [14]$$

com $\Delta y_{it} \equiv y_{it} - y_{it-1}$, $\Delta x_{it} \equiv x_{it} - x_{it-1}$ e $\Delta u_{it} \equiv u_{it} - u_{it-1}$. Tal procedimento elimina a heterogeneidade não observada, uma vez que $\Delta \eta_i = 0$. Essa transformação, conhecida

como primeira diferença, classifica-se como um procedimento do tipo EF e, portanto, dispensa qualquer suposição acerca da correlação entre η_i e x_{it} . Outras transformações capazes de

eliminar o componente não observável η_i também são possíveis nesse contexto, a exemplo da transformação por desvios ortogonais, descrita por Arellano (2003).

Depois de eliminar a heterogeneidade não observada, o procedimento estima os parâmetros em (14) por GMM explorando as suposições de exogeneidade assumidas pelo pesquisador. Se houver, por exemplo, motivo para acreditar que existem efeitos de *feedback* significativos de x para y , não se poderá assumir que x é estritamente exógeno, uma vez que haverá correlação entre u_{it} e $x_{it+1}, x_{it+2}, \dots, x_{iT}$ (isto é, os erros influenciarão os valores futuros de x). Não obstante, se for razoável admitir que não há problemas de simultaneidade, variáveis omitidas (além das capturadas por η_i) ou erros de mensuração que causem correlação entre u_{it} e valores contemporâneos e passados de x , pode-se assumir que esse regressor é sequencialmente exógeno. Mais especificamente, nesse caso, diz-se que x é uma variável “predeterminada” (Arellano, 2003). Sob tal premissa, o estimador pode explorar as seguintes condições de ortogonalidade (ou não correlação), genericamente denominadas condições de momento:

$$E(x_{it-s}\Delta u_{it}) = 0, s \geq 1 \quad [15]$$

Assumindo, sem perda de generalidade, que $E(u_{it}) = 0$, a expressão acima simplesmente reflete a premissa de não correlação entre u_{it} e $x_{it}, x_{it-1}, \dots, x_{i1}$, porém usando os erros transformados Δu_{it} , uma vez que serão estimados os parâmetros do modelo (14), do qual foi removida a heterogeneidade não observada.

Se, no entanto, além dos efeitos de *feedback*, houver, por exemplo, simultaneidade na relação entre x e y , a premissa de não correlação contemporânea entre u e x será violada e a premissa (15), inadequada. Nesse caso, no jargão econométrico, x será uma variável “endógena” e não mais predeterminada. Ainda assim, x não será completamente endógena na medida em que suas defasagens não se correlacionem com o erro do modelo. Em outras palavras, apesar do jargão, x ainda pode ser sequencialmente exógena e, nesse

caso, o estimador GMM-Dif poderá explorar as seguintes condições de momento:

$$E(x_{it-s}\Delta u_{it}) = 0, s \geq 2 \quad [16]$$

Na prática, essas condições de ortogonalidade significam que o estimador utilizará todas as defasagens de x como variáveis instrumentais. A partir daí, seguindo um procedimento análogo ao descrito na seção 2.2 (embora mais complexo do que ele), estima-se o coeficiente de interesse β .

Muitas condições de momento diferentes das representadas por (15) e (16) podem também ser naturalmente acomodadas pelo estimador GMM-Dif, o qual permite o uso como instrumentos não apenas de quaisquer valores passados e futuros de x como também de variáveis externas ao modelo que atendam aos pressupostos descritos na seção 2.2. Naturalmente, no caso particular em que a única fonte de endogeneidade relevante é a presença de heterogeneidade não observada, o GMM-Dif utilizará x como instrumento para si mesmo.

Blundell e Bond (1998) ofereceram a versão final de uma importante extensão do GMM em Diferenças, conhecido como GMM Sistêmico (GMM-Sis). Esse último método aproveita as mesmas condições de momento descritas acima e acrescenta outras, aumentando dessa forma a eficiência e o desempenho em amostras finitas do estimador (Blundell, Bond & Windmeijer, 2000). Continuando o exemplo anterior, se a condição (16) for válida, as seguintes condições de momento adicionais poderão ser exploradas pelo estimador sistêmico:

$$E[\Delta x_{it-1}(\eta_i + u_{it})] = 0 \quad [17]$$

Ao contrário do que se observa em (16), a transformação de primeira diferença é aplicada aos regressores, os quais multiplicam o erro não transformado. Observa-se que esse método impõe a premissa adicional de não correlação entre Δx_{it-1} (ou, mais genericamente, Δx_{it}) e

η_i . Esse último pressuposto não é tão restritivo quanto parece porque permite a correlação entre os regressores e a heterogeneidade não observada. Exige-se apenas que a forma dessa correlação não mude entre um determinado momento do tempo e o momento seguinte, algo frequentemente aceitável, dada a natureza do efeito específico η_i :

$$E(\Delta x_{it}\eta_i) = 0 \Rightarrow E(x_{it}\eta_i) = E(x_{it-1}\eta_i). \quad [18]$$

Blundell e Bond (1998) mostram que a não correlação entre Δx_{it} e η_i será assegurada se o processo estocástico que gera x_{it} for estacionário. Essa é uma condição suficiente e que pode ser testada, mas não é necessária. Condições suficientes mais fracas, relativas ao comportamento dos valores iniciais das séries

$$y_{it} = \alpha + \beta x_{it} + \eta_i + \lambda_t + v_{it}, i = 1, \dots, N, t = 1, \dots, T. \quad [19]$$

Agora, o termo de erro original é decomposto em três componentes: $\varepsilon_{it} = \eta_i + \lambda_t + v_{it}$, sendo η_i a heterogeneidade não observada e v_{it} o termo de erro aleatório. A novidade em (19) é λ_t , que representa os chamados efeitos fixos do tempo. Esse componente varia apenas no tempo e não entre empresas, capturando todo e qualquer choque em y que afetou simultaneamente todas as empresas da amostra.

Não é difícil mostrar que a modelagem explícita de λ_t , frequentemente ignorada, pode ser bastante importante nas pesquisas empíricas em finanças corporativas. Praticamente qualquer variável de resposta de interesse nessa área pode ser significativamente afetada por choques macroeconômicos, a exemplo de variações inesperadas das taxas de inflação e de juros, mudanças do regime cambial ou variações significativas da política fiscal do país. Até certo ponto, por exemplo, todas (ou quase todas) as empresas terão seu desempenho afetado negativamente se houver uma elevação repentina da taxa básica de juros, provocando encarecimento do crédito e redução da demanda. Assim, se y representa desempenho financeiro,

de tempo (x_{i1} , no exemplo) são discutidas por Blundell e Bond (1998, 2000) e Bond (2002).

Em suma, os procedimentos mais avançados de estimação para painel baseados no GMM permitem ao pesquisador recorrer a suposições menos restritivas do que as necessárias para assegurar a consistência dos estimadores tradicionalmente utilizados nas pesquisas empíricas em finanças corporativas. Ademais, eles são particularmente úteis quando o pesquisador não dispõe de variáveis instrumentais externas ao modelo e/ou de contextos quase-experimentais.

3.4 Efeitos fixos no tempo

Uma segunda extensão do modelo básico que relaciona x e y possibilitada pela disposição dos dados em painel é:

o componente comum do choque negativo ocasionado pela elevação dos juros será capturado por λ_t . Na verdade, λ_t captura o impacto sobre y (comum a todas as empresas da amostra) de um conjunto potencialmente amplo de choques macroeconômicos ocorridos no período t (ao longo de um ano, por exemplo). Ainda que os mesmos choques macroeconômicos não exerçam qualquer influência sobre x , ignorar o componente λ_t (portanto, deixando-o dentro do termo de erro do modelo) pode afetar adversamente a estimação dos erros-padrão dos coeficientes (Fama & French, 2002). O problema será maior, entretanto, se λ_t se correlacionar com λ_t . Nesse caso, x será uma variável omitida, tornando inconsistentes os estimadores típicos para dados em painel (incluindo todos aqueles mencionados anteriormente). Isso provavelmente acontecerá se x representar, por exemplo, o porte da empresa (medido por seu faturamento líquido), seu nível de endividamento, lucro ou mesmo a participação acionária dos gestores.

Felizmente, é perfeitamente viável isolar o impacto potencialmente relevante de λ_t e a forma mais prática de fazê-lo é incluir na regressão um

conjunto de variáveis indicadoras de tempo (d_t , sendo $t = 1, \dots, T$) de forma que $d_t = 1$ no período t e $d_t = 0$, caso contrário (naturalmente, essa variável dispensa o subscrito i porque não varia entre empresas). Logo, o modelo efetivamente

$$y_{it} = \alpha + \beta x_{it} + \eta_i + \sum_{t=2}^T d_t + v_{it}, i = 1, \dots, N, t = 1, \dots, T. \quad [20]$$

3.5 Modelos dinâmicos

Os modelos formulados até aqui desconsideram a possível influência direta de valores passados da variável de resposta sobre seus valores contemporâneos. Boa parte dos indicadores de interesse para finanças, no entanto, apresenta comportamento fortemente inercial (e.g., práticas de governança, desempenho financeiro, alavancagem, faturamento), sugerindo que a especificação de modelos estáticos pode não ser adequada.

Diferentes argumentos podem justificar tal comportamento. Wintoki et al. (2012), por exemplo, sugerem que a elevada persistência temporal da lucratividade das empresas, verificada em diversos estudos empíricos (Glen, Lee & Singh, 2001; Waring, 1996), reflete, em alguma medida,

estimado (por qualquer dos métodos discutidos anteriormente) será (excluiu-se d_1 da equação para evitar colinearidade perfeita dos regressores, uma vez que o modelo inclui intercepto):

variáveis não observadas como a habilidade gerencial (a qual pode ter alguma variação no tempo, não sendo por isso perfeitamente capturada pelo efeito fixo). Ademais, é comum observar-se algum comportamento de regressão à média em variáveis corporativas, induzindo uma correlação negativa entre os valores atuais dessas variáveis e suas variações subsequentes. De fato, tal movimento de ajustamento parcial em direção a valores de equilíbrio é esperado, por exemplo, por diferentes teorias de estrutura de capital que sugerem a existência de uma estrutura ótima de financiamentos para cada empresa (Fama & French, 2002; Frank & Goyal, 2003).

Para modelar explicitamente o componente dinâmico do modelo que relaciona as variáveis de interesse y e x , podemos, por exemplo, estender (19) na forma:

$$y_{it} = \alpha_0 + \alpha_1 y_{it-1} + \beta x_{it} + \eta_i + \lambda_t + v_{it}, i = 1, \dots, N, t = 1, \dots, T. \quad [21]$$

Se o modelo correto for representado por (21), sendo $\alpha_1 \neq 0$, a omissão de y_{it-1} na regressão enviesará o estimador de β na medida em que y_{it-1} (que comporá o termo de erro do modelo estimado) se correlacione com x_{it} . Uma condição suficiente para que isso ocorra é que x tenha alguma persistência temporal, de forma que exista correlação significativa entre x_{it} e x_{it-1} . Naturalmente, uma fonte ainda mais direta de viés, nesse caso, seria a existência de retroalimentação de y para x , como discutido na seção 3.2.

Um indício de inadequação da especificação estática é a presença de significativa autocorrelação em v_{it} , o que pode ser verificado empiricamente pelo pesquisador por meio de testes de autocorrelação dos resíduos da regressão estática original. Em muitos casos, a inclusão da primeira defasagem da variável de resposta entre os regressores é suficiente para capturar esse fenômeno, mas, em tese, outras defasagens também podem ser relevantes para dar conta do comportamento dinâmico de y (e.g., $y_{it-1}, y_{it-2}, \dots, y_{it-p}$).

O modelo (21) não será adequadamente estimado por qualquer procedimento que necessite da suposição de exogeneidade estrita dos regressores, como é o caso dos tradicionais estimadores de EF e EA, uma vez que, por definição, y_{it-1} não é uma variável estritamente exógena. Tal pressuposto, nesse modelo, implicaria a não correlação entre v_{it} e y observado em qualquer instante do tempo. Portanto, ele exigiria também a não correlação entre v_{it} e y_{it} , algo impossível por construção. Não obstante, se os regressores forem sequencialmente exógenos, os parâmetros de (21) podem ser consistentemente estimados pelos métodos baseados no GMM apresentados na seção 3.3.

4 Resultados das regressões com dados simulados e desempenho dos estimadores

Nesta seção, são apresentados os procedimentos para construção de amostras em painel simuladas com características similares às disponíveis para os pesquisadores de finanças corporativas. Em seguida, apresentamos alguns resultados de regressões empregando estimadores mais simples e mais avançados aplicados às amostras simuladas, permitindo a comparação

de seu desempenho relativo e a avaliação da adequação das diferentes estratégias de estimação aos dados gerados. Na medida em que esta análise laboratorial sintetiza aspectos salientes e importantes dos dados tipicamente utilizados nas pesquisas empíricas de finanças corporativas, ela pode oferecer um direcionador metodológico para os estudiosos da área, por um lado apontando os problemas mais graves aos quais eles deveriam atentar e, por outro, oferecendo possíveis soluções.

4.1 Modelo geral da simulação

O objetivo desta análise de simulação é avaliar o desempenho de diferentes estratégias de estimação aplicadas a amostras de dados artificiais com características similares às efetivamente utilizadas pelos pesquisadores de finanças corporativas em seus estudos empíricos. Para tanto, geramos, por meio de procedimentos de Monte Carlo, conjuntos de amostras aleatórias baseadas em modelos que sintetizam as referidas características da forma mais completa possível.

O modelo geral da simulação é bastante similar ao mostrado em (21) (excluindo, por simplicidade e sem perda de generalidade, o intercepto α_0), sendo v_{it} seu termo de erro aleatório:

$$y_{it} = \alpha y_{it-1} + \beta x_{it} + \eta_i + \lambda_t + v_{it}, i = 1, \dots, N, t = 1, \dots, T. \quad [22]$$

Ele captura diversas características potencialmente relevantes de processos de interesse para pesquisadores de finanças corporativas, incluindo o comportamento dinâmico da variável de resposta (representado por y_{it-1}), a heterogeneidade não observada das empresas (η_i) e a influência de fatores macroeconômicos não observados⁴ (λ_t).

Tão importante quanto modelar o comportamento da variável de resposta, não obstante, é modelar o comportamento do regressor de interesse x , para que se contemple na análise diversos problemas de endogeneidade capazes de impedir a estimação consistente dos parâmetros de interesse α e β . O modelo geral para x é mostrado a seguir (sendo e_{it} seu termo de erro aleatório):

$$x_{it} = \rho x_{it-1} + \tau \eta_i + \phi \lambda_t + \theta_1 v_{it} + \theta_2 v_{it-1} + e_{it}, i = 1, \dots, N, t = 1, \dots, T. \quad [23]$$

O modelo (23) também permite que x exiba algum grau de persistência temporal (como

parece ser o caso da maior parte das variáveis corporativas) e permite contemplar todos os

problemas de endogeneidade discutidos nas seções anteriores, como explicamos a seguir.

O problema de variáveis omitidas relacionado a características não observáveis e invariantes no tempo das empresas da amostra é representado por $\tau\eta_i$ e existirá se $\tau \neq 0$, sendo tanto maior quanto maior for esse parâmetro. Similarmente, o problema de variáveis omitidas relacionado a efeitos do tempo (e.g., choques macroeconômicos) não observáveis é representado por $\phi\lambda_t$ e será proporcional ao valor do parâmetro ϕ . Por sua vez, a endogeneidade de x motivada por efeitos de retroalimentação de y para x (também chamada de endogeneidade dinâmica, conforme discutido na seção 3.2) é capturada por θ_2v_{it-1} (Y_{it-1} poderia ser utilizado no lugar de v_{it-1} com resultados similares) e sua magnitude dependerá do valor associado a θ_2 . A possível (e provável, em muitos contextos empíricos de finanças corporativas) determinação simultânea de y e x é capturada por θ_1v_{it} , uma vez que o fenômeno da causalidade reversa produzirá correlação contemporânea entre v e x . Por fim, tanto θ_1v_{it} quanto θ_2v_{it-1} servem também para capturar a

possível endogeneidade motivada por erros de mensuração em x ou y , além de variáveis omitidas que variam no tempo e entre empresas.

A construção de amostras simuladas baseadas nos modelos (22) e (23) permite analisar com precisão os efeitos combinados de diferentes problemas de endogeneidade aplicáveis a pesquisas empíricas no campo de finanças corporativas. Como resultado, esse exercício computacional permite destacar os desafios mais críticos para a estimação consistente dos parâmetros de interesse em regressões com dados observacionais, bem como apontar algumas estratégias para enfrentá-los. Para cumprir a contento esses objetivos, todavia, é também importante reduzir o modelo geral a casos particulares significativos, dessa forma isolando problemas específicos, como será discutido nas seções seguintes.

O modelo completo utilizado inicialmente para gerar os painéis simulados desta pesquisa é apresentado abaixo de forma mais detalhada, incluindo os parâmetros escolhidos, a título de ilustração, pelos pesquisadores.

$$\begin{aligned} y_{it} &= \alpha y_{it-1} + \beta x_{it} + \eta_i + \lambda_t + v_{it} \\ x_{it} &= \rho x_{it-1} + \tau \eta_i + \phi \lambda_t + \theta_1 v_{it} + \theta_2 v_{it-1} + e_{it}, \end{aligned} \quad [24]$$

tal que $\eta_i \sim N(0, \sigma_\eta^2)$, $\lambda_t \sim N(0, \sigma_\lambda^2)$, $v_{it} \sim N(0, \sigma_v^2)$, $e_{it} \sim N(0, \sigma_e^2)$. Nesta pesquisa, assumimos: $\sigma_\eta^2 = \sigma_\lambda^2 = \sigma_v^2 = \sigma_e^2 = 1$, $\beta = 1$, $\tau = 0.7$, $\alpha = 0.5$, $\theta_1 = 0.6$, $\theta_2 = 0.5$, $\rho = 0.5$ e $\phi = 0.5$.

Como mostrado em (24), assumimos que η_i , λ_t , v_{it} e e_{it} são variáveis aleatórias que seguem uma distribuição normal padrão univariada. Essa escolha, todavia, não implica perda de generalidade, uma vez que os procedimentos de estimação utilizados a seguir são assintoticamente robustos a desvios da normalidade. O código de programação original foi desenvolvido para Matlab e utilizado para gerar as amostras conforme o sistema de equações (24).

4.1.1 Desempenho dos estimadores com base no modelo geral da simulação

Depois de gerados os dados com base no modelo (24), com $N = 500$, $T = 8$ e 1000

replicações, estimamos, para cada uma das 1000 amostras, os parâmetros de interesse α e β com base em cinco diferentes métodos de estimação. Especificamente, utilizamos o tradicional estimador MQO, os estimadores de EA e EF, bem como os métodos baseados no GMM (GMM-Dif e GMM-Sis). A implementação de todos os procedimentos de estimação foi feita no pacote estatístico Stata, utilizando o a função 'xtabond2'. Todos os códigos para Matlab e Stata serão disponibilizados pelos autores a pedido.

Os resultados da estimação do modelo geral são reportados na Tabela 1. Embora o modelo descrito pelo sistema de equações (24) seja afetado por diferentes problemas de

endogeneidade, a estimação por MQO só é capaz de evitar o viés de variáveis omitidas motivado por λ_t ao incluir entre os regressores variáveis indicadoras (*dummies*) de tempo (vide seção 3.4). Os demais problemas são forçosamente ignorados e isso resulta em substancial enviesamento do estimador de β , considerando-se que o valor verdadeiro do parâmetro é 1 e o valor médio das 1000 estimativas computadas é igual a 1.3678 (com mínimo de 1.3292 e máximo de 1.4113). A distância entre o valor verdadeiro e os obtidos pelo estimador se reflete também no elevado erro quadrático médio (RMSE), associado ao parâmetro β . O RMSE de β é calculado pela seguinte equação:

$$RMSE = \frac{\sum_{j=1}^S (\hat{\beta}_j - \beta)^2}{S}, \quad [25]$$

tal que $\hat{\beta}_j$ é a estimativa desse parâmetro na j -ésima amostra simulada (de um total de S amostras). Na Tabela 1, $S = 1000$ e $\beta = 1$.

Um viés dessa magnitude seria economicamente relevante se os dados correspondessem a informações financeiras de empresas reais. Com base na estimativa média mostrada na Tabela 1, por exemplo, um pesquisador que utilizasse o estimador MQO poderia inferir que, mantendo constantes outros fatores, o aumento de x em uma unidade ocasionaria um aumento esperado de aproximadamente 1.4

unidade em y , uma distorção de 40% em relação à inferência correta. Naturalmente, os valores aqui apresentados são apenas ilustrativos. Na vida real, o tamanho do problema dependerá de diversos fatores, incluindo variação amostral e a magnitude da correlação entre os regressores e o termo de erro do modelo. A Tabela 1 também mostra um enviesamento do estimador de α , embora menos pronunciado, quando se utiliza a estimação via MQO.

O modelo seguinte, com estimação por EA, produz resultados com enviesamento similar ao reportado para o estimador MQO. Embora esse procedimento inclua explicitamente a heterogeneidade não observada (η_i), assume que ela não se correlaciona com os regressores. Ademais, esse procedimento não é capaz de lidar com outras fontes de endogeneidade, como efeitos de *feedback*, erros de mensuração dos regressores e simultaneidade.

O resultado do modelo seguinte mostra também um enviesamento substancial do estimador por EF. Embora esse procedimento seja mais robusto que os anteriores, permitindo a livre correlação entre η_i e os regressores, sua validade depende fundamentalmente do pressuposto de que os regressores são estritamente exógenos. A violação dessa premissa no modelo (24), combinada com os demais problemas de endogeneidade, resulta em um acentuado enviesamento de $\hat{\alpha}$ (maior do que com os métodos anteriores), cujo valor verdadeiro é 0.5, bem como de $\hat{\beta}$.

Tabela 1

Resumo das estimações para o modelo geral

Modelo	Parâm.	Média	DP	Mín.	Máx.	RMSE
MQO	α	0,5344	0,0055	0,5152	0,5477	0,0349
	β	1,3678	0,0126	1,3292	1,4113	0,3680
EA	α	0,5167	0,0064	0,4964	0,5352	0,0179
	β	1,3802	0,0124	1,3429	1,4208	0,3804
EF	α	0,3134	0,0077	0,2835	0,3398	0,1868
	β	1,3510	0,0119	1,3158	1,3845	0,3512
GMM-Dif	α	0,4755	0,0239	0,3941	0,5481	0,0342
	β	1,0440	0,0426	0,9177	1,1645	0,0612
GMM-Sis	α	0,4908	0,0228	0,4269	0,5654	0,0245
	β	1,0356	0,0438	0,8640	1,1561	0,0564

Nota: Regressões baseadas no sistema de equações (24), com parâmetros populacionais: $\beta = 1, \tau = 0,7, \alpha = 0,5, \theta_1 = 0,6, \theta_2 = 0,5, \rho = 0,5$ e $\phi = 0,5$. Número de amostras simuladas: 1000. Parâm.: Parâmetros estimados; DP: Desvio-padrão; RMSE: *Root Mean Square Error*. Todas as regressões incluem *dummies* de tempo e variável dependente defasada. Estimadores: Mínimos Quadrados Ordinários (MQO); Efeitos Aleatórios (EA); Efeitos Fixos (EF); Método dos Momentos Generalizado (GMM) em Diferenças (GMM-Dif); GMM Sistemático (GMM-Sis).

O estimador GMM-Dif consegue lidar apropriadamente com todas as fontes de endogeneidade incluídas no sistema de equações (24), removendo a heterogeneidade não observada e utilizando defasagens de y e de x não correlacionadas com o erro v_{it} como variáveis instrumentais. Observa-se que as estimativas de β ficam bem mais próximas de seu valor verdadeiro quando comparadas com os métodos anteriores. O resultado para $\hat{\alpha}$, entretanto, é menos satisfatório, uma vez que suas estimativas, embora próximas do valor verdadeiro de 0.5 em média, variam entre 0.3941 e 0.5481.

O modelo seguinte revela que o método GMM-Sis produz os resultados mais satisfatórios dentre todos os empregados, com enviesamento quase nulo tanto para $\hat{\alpha}$ quanto para $\hat{\beta}$ e baixo RMSE. Sua vantagem em comparação com o GMM-Dif é o emprego de instrumentos adicionais baseados em premissas de exogeneidade sequencial dos regressores.

4.2 Caso particular 1: correlação entre o regressor e a heterogeneidade não observada

Além de analisar o modelo geral, esta pesquisa investigou o comportamento de diversos casos particulares, isto é, reduções do modelo

geral que permitissem investigar o desempenho dos estimadores em contextos mais específicos, isolando os problemas possíveis de endogeneidade encontrados em dados reais.

O caso particular 1 enfoca apenas o problema da heterogeneidade não observada correlacionada com x , eliminando as demais fontes potenciais de endogeneidade. Para tanto, é necessário gerar toda uma nova análise de simulação a partir da alteração dos parâmetros do modelo geral, isto é, associando a alguns deles o valor zero. Como o caso particular 1 se baseia em um modelo estático, por exemplo, fazemos $\alpha = 0$. Portanto, o modelo geral agora se reduz a:

$$\begin{aligned} y_{it} &= \beta x_{it} + \eta_i + v_{it} \\ x_{it} &= \rho x_{it-1} + \tau \eta_i + e_{it} \end{aligned} \quad [26]$$

A diferença entre os modelos apresentados nas equações (26) e (24) é que a única fonte de endogeneidade do primeiro é correlação entre η_i e x_{it} . Portanto, (26) é um modelo bem mais simples. De resto, sua especificação é idêntica à do modelo geral (24). Um raciocínio análogo se aplica aos demais casos particulares. A Tabela 2 reporta os resultados desta análise.

Tabela 2

Resumo das estimações para o caso particular 1

Modelo	Parâm.	Média	SD	Mín.	Máx.	RMSE
MQO	β	1,4248	0,0147	1,3775	1,4672	0,4251
EA	β	1,3059	0,0182	1,2493	1,3565	0,3064
EF	β	1,0007	0,0164	0,9508	1,0563	0,0163
GMM-Dif	β	1,0006	0,0222	0,9159	1,0659	0,0222

Nota: Regressões baseadas no sistema de equações (26), com parâmetros populacionais: $\beta = 1, \tau = 0,7$ e $\rho = 0,5$. Número de amostras simuladas: 1000. Parâm.: Parâmetros estimados; DP: Desvio-padrão; RMSE: *Root Mean Square Error*. Regressões simples (x é o único regressor). Estimadores: Mínimos Quadrados Ordinários (MQO); Efeitos Aleatórios (EA); Efeitos Fixos (EF); Método dos Momentos Generalizado (GMM) em Diferenças (GMM-Dif).

Há um enviesamento substancial dos estimadores nos modelos estimados por MQO e EA, principalmente do primeiro, decorrente de sua incapacidade de controlar a endogeneidade existente. Por outro lado, os estimadores baseados em efeitos fixos (EF e GMM-Dif) apresentam resultados bastante satisfatórios, como esperado. De fato, o estimador EF, nesse caso, apresentou o melhor desempenho de todos, marginalmente superior ao do GMM-Dif. A tabela omite as estimativas via GMM-Sis porque, nesse modelo simples e estático, os dois estimadores GMM produzem resultados praticamente indistinguíveis.

4.3 Caso particular 2: persistência temporal da variável resposta

O caso particular 2 enfoca a importância de se incluir no modelo termos dinâmicos quando existe persistência temporal da variável

de resposta. É comum, em estudos empíricos de finanças corporativas, a especificação exclusiva de modelos estáticos ($\alpha = 0$) e a Tabela 3 ilustra as consequências dessa especificação potencialmente inadequada do modelo empírico.

$$\begin{aligned} y_{it} &= \alpha y_{it-1} + \beta x_{it} + v_{it} \\ x_{it} &= \rho x_{it-1} + e_{it}. \end{aligned} \quad [27]$$

A Tabela 3 mostra um enviesamento substancial de todos os estimadores quando não incluímos y_{it-1} entre os regressores do modelo pelo sistema de equações (27), principalmente o estimador MQO. É interessante notar, também, que o GMM-Dif tende a subestimar o parâmetro β , ao passo que os demais tendem a superestimá-lo. As estimativas produzidas por EA e GMM-Sis são omitidas por economia de espaço.

Tabela 3

Resumo das estimações para o caso particular 2

Modelo	Parâm.	Média	SD	Mín.	Máx.	RMSE
MQO-Est	β	1,3330	0,0247	1,2663	1,4063	0,3339
MQO-Din	α	0,4998	0,0088	0,4725	0,5263	0,0088
	β	1,0003	0,0158	0,9493	1,0463	0,0158
EF-Est	β	1,1042	0,0230	1,0390	1,1684	0,1067
	α	0,4064	0,0118	0,3697	0,4434	0,0943
EF-Din	β	1,0144	0,0183	0,9594	1,0704	0,0232
	α	0,4981	0,0178	0,4483	0,5643	0,0179
GMM-Dif-Est	β	0,8329	0,0182	0,7659	0,8911	0,1681
GMM-Dif-Din	β	0,9957	0,0229	0,9187	1,0685	0,0233

Nota: Regressões baseadas no sistema de equações (27), com parâmetros populacionais: $\beta = 1, \alpha = 0,5$ e $\rho = 0,5$. Número de amostras simuladas: 1000. Parâm.: Parâmetros estimados; DP: Desvio-padrão; RMSE: *Root Mean Square Error*. Nas especificações dinâmicas (Din) inclui-se a variável dependente defasada, a qual é excluída das especificações estáticas (Est). Estimadores: Mínimos Quadrados Ordinários (MQO); Efeitos Fixos (EF); Método dos Momentos Generalizado (GMM) em Diferenças (GMM-Dif).

4.4 Caso particular 3: efeitos de retroalimentação

O caso particular 3 avalia a importância dos efeitos de retroalimentação de y para x , capturados pelo termo $\theta_2 v_{it-1}$. A Tabela 4 apresenta os resultados das estimações.

$$\begin{aligned} y_{it} &= \alpha y_{it-1} + \beta x_{it} + v_{it} \\ x_{it} &= \rho x_{it-1} + \theta_2 v_{it-1} + e_{it}. \end{aligned} \quad [28]$$

Como esperado, o estimador MQO estima os parâmetros adequadamente nesse caso, uma vez que não depende da premissa de

exogeneidade estrita e, portanto, não é afetado pelo fenômeno da endogeneidade dinâmica. O mesmo não ocorre, entretanto, com o estimador EF, que depende da premissa de exogeneidade estrita. Uma vez que no sistema de equações (28) a referida premissa é violada, os coeficientes são estimados de forma inconsistente e a análise mostra que o parâmetro $\hat{\alpha}$ é o mais afetado pelo problema. Por sua vez, os estimadores GMM-Dif e GMM-Sis, adotando a premissa de que x é uma variável predeterminada, apresentam novamente resultados bastante satisfatórios, com vantagem marginal para o último estimador. As estimativas via EA são omitidas por economia de espaço.

Tabela 4
Resumo das estimações para o caso particular 3

Modelo	Parâm.	Média	SD	Mín.	Máx.	RMSE
MQO	α	0,4997	0,0083	0,4751	0,5240	0,0083
	β	1,0001	0,0157	0,9500	1,0470	0,0157
EF	α	0,3820	0,0108	0,3479	0,4190	0,1185
	β	0,9710	0,0173	0,9240	1,0286	0,0337
GMM-Dif	α	0,4955	0,0148	0,4428	0,5411	0,0154
	β	0,9966	0,0235	0,9182	1,0740	0,0238
GMM-Sis	α	0,4992	0,0124	0,4614	0,5317	0,0124
	β	1,0001	0,0180	0,9418	1,0570	0,0180

Nota: Regressões baseadas no sistema de equações (28), com parâmetros populacionais: $\beta = 1, \alpha = 0,5, \theta_2 = 0,5$ e $\rho = 0,5$. Número de amostras simuladas: 1000. Parâm.: Parâmetros estimados; DP: Desvio-padrão; RMSE: *Root Mean Square Error*. Todas as regressões incluem a variável dependente defasada. Estimadores: Mínimos Quadrados Ordinários (MQO); Efeitos Fixos (EF); Método dos Momentos Generalizado (GMM) em Diferenças (GMM-Dif); GMM Sistemico (GMM-Sis).

4.5 Caso particular 4: heterogeneidade não observada e retroalimentação

O caso particular 4 difere do caso 3 em dois aspectos: ele inclui no modelo a heterogeneidade não observada, mantendo os efeitos de retroalimentação de y para x , e retira o termo dinâmico. Portanto, agora x será uma variável predeterminada e correlacionada com o efeito fixo η_i . Logo, existirão duas fontes de endogeneidade simultaneamente. Os resultados das simulações estão apresentados na Tabela 5.

$$\begin{aligned} y_{it} &= \beta x_{it} + \eta_i + v_{it} \\ x_{it} &= \rho x_{it-1} + \tau \eta_i + \theta_2 v_{it-1} + e_{it} \end{aligned} \quad [29]$$

A Tabela 5 mostra claramente o substancial enviesamento para cima dos estimadores MQO (por causa de $\tau \eta_i$) e EA (causado pela interação entre $\tau \eta_i$ e $\theta_2 v_{it-1}$), sendo significativamente maior o enviesamento do primeiro. Embora em menor grau, também se mostra substancialmente enviesado (dessa vez para baixo, ao contrário dos anteriores) o estimador EF em razão do efeito de *feedback* relacionado a $\theta_2 v_{it-1}$. Apenas os

estimadores GMM-Dif e GMM-Sis conseguem estimar consistentemente o parâmetro β , com

vantagem marginal para o segundo método, que se mostra mais eficiente.

Tabela 5

Resumo das estimações para o caso particular 4

Modelo	Parâm.	Média	SD	Mín.	Máx.	RMSE
MQO	β	1,3851	0,0144	1,3407	1,4439	0,3854
EA	β	1,2889	0,0186	1,2376	1,3457	0,2894
EF	β	0,9178	0,0146	0,8598	0,9678	0,0835
GMM-Dif	β	0,9971	0,0264	0,9080	1,0948	0,0265
GMM-Sis	β	1,0040	0,0226	0,9273	1,0653	0,0229

Nota: Regressões baseadas no sistema de equações (29), com parâmetros populacionais: $\beta = 1, \tau = 0,7, \theta_2 = 0,5$ e $\rho = 0,5$. Número de amostras simuladas: 1000. Parâm.: Parâmetros estimados; DP: Desvio-padrão; RMSE: *Root Mean Square Error*. Regressões simples (x é o único regressor). Estimadores: Mínimos Quadrados Ordinário (MQO); Efeitos Aleatórios (EA); Efeitos Fixos (EF); Método dos Momentos Generalizado (GMM) em Diferenças (GMM-Dif); GMM Sistemico (GMM-Sis).

4.6 Caso particular 5: heterogeneidade não observada, erros de mensuração e/ou determinação simultânea

O caso particular 5 é similar ao caso 4, mas agora modelamos x como uma variável contemporaneamente correlacionada com v , em razão, por exemplo, de erros de mensuração e/ou

de sua determinação simultânea com a variável de resposta, problema frequente em estudos com dados de empresas. Os resultados são apresentados na Tabela 6.

$$\begin{aligned}
 y_{it} &= \beta x_{it} + \eta_i + v_{it} \\
 x_{it} &= \rho x_{it-1} + \tau \eta_i + \theta_1 v_{it} + e_{it}.
 \end{aligned}
 \tag{30}$$

Tabela 6

Resumo das estimações para o caso particular 5

Modelo	Parâm.	Média	SD	Mín.	Máx.	RMSE
MQO	β	1,5301	0,0109	1,4945	1,5659	0,5303
EA	β	1,4929	0,0111	1,4523	1,5230	0,4930
EF	β	1,3768	0,0126	1,3370	1,4187	0,3770
GMM-Dif	β	1,0285	0,0561	0,8353	1,1919	0,0629
GMM-Sis	β	1,0381	0,0476	0,8715	1,1782	0,0609

Nota: Regressões baseadas no sistema de equações (30), com parâmetros populacionais: $\beta = 1, \tau = 0,7, \theta_1 = 0,6$ e $\rho = 0,5$. Número de amostras simuladas: 1000. Parâm.: Parâmetros estimados; DP: Desvio-padrão; RMSE: *Root Mean Square Error*. Regressões simples (x é o único regressor). Estimadores: Mínimos Quadrados Ordinários (MQO); Efeitos Aleatórios (EA); Efeitos Fixos (EF); Método dos Momentos Generalizado (GMM) em Diferenças (GMM-Dif); GMM Sistemico (GMM-Sis).

Verifica-se agora um enviesamento bem maior do que no caso anterior dos estimadores MQO, EA e EF, ilustrando o impacto substancial de problemas de endogeneidade causados, por exemplo, por questões de causalidade reversa entre os regressores e a variável resposta. Uma vez que, além desse tipo de endogeneidade,

existe correlação entre x e η_i , não é surpresa que o viés maior seja dos estimadores MQO e EA. Mais uma vez, os estimadores GMM-Dif e GMM-Sis produziram bons resultados, com viés insignificante e alguma vantagem em termos de precisão para o estimador sistemico (GMM-Sis).



4.7 Caso particular 6: erros de mensuração e/ou determinação simultânea sem heterogeneidade não observada

O caso particular 6 difere do caso anterior por focar apenas a correlação contemporânea entre x e v , removendo do sistema de equações

(31) a heterogeneidade não observada. Os resultados são apresentados na Tabela 7.

$$\begin{aligned} y_{it} &= \beta x_{it} + v_{it} \\ x_{it} &= \rho x_{it-1} + \theta_1 v_{it} + e_{it} \end{aligned} \quad [31]$$

Tabela 7

Resumo das estimações para o caso particular 6

Modelo	Parâm.	Média	SD	Mín.	Máx.	RMSE
MQO	β	1,3314	0,0110	1,2918	1,3616	0,3316
EA	β	1,3316	0,0110	1,2925	1,3616	0,3318
EF	β	1,3765	0,0133	1,3396	1,4266	0,3767
GMM-Dif	β	1,0164	0,0435	0,8759	1,1350	0,0465
GMM-Sis	β	1,0180	0,0360	0,8701	1,1540	0,0403

Nota: Regressões baseadas no sistema de equações (31), com parâmetros populacionais: $\beta = 1, \theta_1 = 0,6$ e $\phi = 0,5$. Número de amostras simuladas: 1000. Param.: Parâmetros estimados; DP: Desvio-padrão; RMSE: *Root Mean Square Error*. Regressões simples (x é o único regressor). Estimadores: Mínimos Quadrados Ordinários (MQO); Efeitos Aleatórios (EA); Efeitos Fixos (EF); Método dos Momentos Generalizado (GMM) em Diferenças (GMM-Dif); GMM Sistemático (GMM-Sis).

Uma vez que a correlação contemporânea entre x e v é a única fonte de endogeneidade nesse caso, o viés dos estimadores MQO, EA e EF é menor do que no caso anterior, mas ainda bastante pronunciado. Também como se esperava, em razão da ausência da heterogeneidade não observada, o grau de enviesamento dos três estimadores é similar. Novamente, os estimadores GMM-Dif e GMM-Sis produziram resultados corretos, com ligeira vantagem em termos de precisão para o estimador sistemático.

4.8 Caso particular 7: efeitos fixos no tempo

No caso particular 7, o foco recai sobre a importância dos efeitos fixos do tempo, capturados por λ_t , e sua possível correlação com os regressores do modelo, representada por $\phi\lambda_t$. Nesse caso, os modelos são estimados com e sem variáveis indicadoras (*dummies*) de tempo, possibilitando avaliar o impacto potencial da omissão desses regressores no sistema de equações (32). Os resultados são apresentados na Tabela 8.

$$\begin{aligned} y_{it} &= \beta x_{it} + \lambda_t + v_{it} \\ x_{it} &= \rho x_{it-1} + \phi\lambda_t + e_{it} \end{aligned} \quad [32]$$

Os resultados da análise de simulação mostram que a omissão das *dummies* de tempo quando existem efeitos macroeconômicos relevantes que afetam x e y simultaneamente pode conduzir a um enviesamento substancial de qualquer dos estimadores empregados. Em particular, observa-se que os estimadores GMM-Dif e GMM-Sis são, respectivamente, os que apresentam maior grau de enviesamento quando se ignora o problema. Esse resultado, talvez surpreendente, alerta para a importância de se incluir *dummies* de tempo em regressões com dados em painel em estudos envolvendo variáveis corporativas, as quais muito provavelmente são influenciadas por choques macroeconômicos ou fenômenos cíclicos similares.

Tabela 8

Resumo das estimações para o caso particular 7

Modelo	Parâm.	D(tempo)	Média	SD	Mín.	Máx.	RMSE
MQO	β	não	1,2418	0,1107	0,9737	1,5901	0,2659
	β	sim	0,9999	0,0133	0,9586	1,0698	0,0134
EF	β	não	1,3268	0,1411	0,9746	1,7401	0,3560
	β	sim	0,9998	0,0157	0,9554	1,0677	0,0157
GMM-Dif	β	não	1,4356	0,2079	0,9877	2,1720	0,4826
	β	sim	0,9999	0,0222	0,9330	1,0651	0,0222
GMM-Sis	β	não	1,3611	0,1497	1,0356	1,8523	0,3908
	β	sim	0,9999	0,0157	0,9496	1,0642	0,0156

Nota: Regressões baseadas no sistema de equações (32), com parâmetros populacionais: $\beta = 1, \rho = 0,5$ e $\phi = 0,5$. Número de amostras simuladas: 1000. Parâm.: Parâmetros estimados; D(tempo): *dummies* de tempo incluídas nas regressões; DP: Desvio-padrão; RMSE: *Root Mean Square Error*. Estimadores: Mínimos Quadrados Ordinários (MQO); Efeitos Aleatórios (EA); Efeitos Fixos (EF); Método dos Momentos Generalizado (GMM) em Diferenças (GMM-Dif); GMM Sistemico (GMM-Sis).

5 Considerações finais

A maior parte dos trabalhos empíricos em finanças corporativas utiliza dados observacionais de empresas com o intuito de discernir relações de causalidade entre variáveis empregando regressões lineares. Em quase todos os estudos dessa área, entretanto, o pesquisador depara com o desafio de identificar e lidar com os diferentes problemas de endogeneidade dos regressores, os quais, se ignorados, podem conduzir a inferências inadequadas. Neste estudo, discutimos as principais causas do problema e suas possíveis soluções, em particular quando o pesquisador dispõe de dados em painel, mas não de variáveis instrumentais externas ao modelo nem de contextos quase-experimentais.

Por meio de amostras simuladas que emulam algumas das principais características dos dados utilizados em finanças corporativas, ilustramos o impacto potencial das variadas formas de endogeneidade, bem como algumas soluções disponíveis, comparando-se a eficácia relativa de diferentes métodos de estimação.

Os resultados mostram claramente o potencial enviesamento dos coeficientes estimados quando não se trata adequadamente dos problemas de variáveis omitidas, erros de mensuração dos regressores, determinação simultânea de variáveis explicativas e explicadas

ou dos efeitos de retroalimentação, também conhecidos como endogeneidade dinâmica. As implicações conjuntas e separadas dessas questões são avaliadas por meio de um modelo geral e de sete casos particulares estimados pelos procedimentos de Mínimos Quadrados Ordinários (MQO), Efeitos Fixos (EF), Efeitos Aleatórios (EA), GMM em Diferenças (GMM-Dif) e GMM Sistemico (GMM-Sis).

As análises revelam que os tradicionais métodos de MQO, EA e EF podem ser inconsistentes na presença de problemas de endogeneidade bastante plausíveis no contexto de finanças corporativas. Por outro lado, os métodos de estimação para dados em painel baseados no GMM que se utilizam de suposições de exogeneidade sequencial dos regressores apresentam-se como alternativas capazes de contornar eficazmente todos os problemas elencados (desde que as premissas adotadas sejam válidas) mesmo que o pesquisador não disponha de boas variáveis instrumentais externas ao modelo. Em particular, as análises de simulação apontam para o estimador GMM-Sis (Blundell & Bond, 1998) como o mais interessante (combinando baixo viés e elevada eficiência) para modelar empiricamente relacionamentos causais entre variáveis corporativas.

Naturalmente, a eficácia desses procedimentos dependerá da validade das

referidas premissas de exogeneidade sequencial e da especificação adequada do modelo empírico, algo que não se pode assegurar *a priori*. A validade das premissas de exogeneidade sequencial, por exemplo, depende da ausência de autocorrelação nos erros do modelo; o pressuposto de estabilidade temporal da correlação entre os regressores e a heterogeneidade não observada (vide equação (18)), crucial para o estimador GMM-Sis, pode não ser realista; se o número de instrumentos derivados das suposições de exogeneidade sequencial for muito elevado em comparação com o número de observações da amostra os estimadores GMM analisados podem ser enviesados, fenômeno conhecido como *overfitting* (Bun & Sarafidis, 2015); Dang et al. (2015) destacam que certas características dos dados podem gerar distorções na estimação, a exemplo da censura dos valores da variável dependente (e.g., restrição dos valores da variável dependente ao intervalo [0,1]); pode haver relevante viés de amostra finita quando os instrumentos utilizados são insuficientemente correlacionados com os regressores instrumentados (os instrumentos são fracos; Windmeijer, 2018, e Wintoki et al., 2012).

Variados testes de diagnóstico estão disponíveis para os estimadores analisados nesta pesquisa (em especial os baseados no GMM), cujo intuito é verificar a adequação do modelo aos pressupostos que asseguram a validade do procedimento de estimação, a exemplo de testes de autocorrelação dos erros e testes de validade das variáveis instrumentais. Embora eles sirvam como importantes guias para a modelagem empírica, infelizmente seu uso não garante que todos os problemas serão detectados. Wintoki et al. (2012) destacam, por exemplo, que os testes de adequação dos instrumentos (i.e., testes de sobreidentificação) frequentemente falham em detectar problemas de endogeneidade das variáveis instrumentais. Outro problema de difícil diagnóstico, até recentemente, era a detecção de enviesamentos causados por instrumentos fracos no contexto da estimação de modelos de painel dinâmico por GMM. Felizmente, o procedimento proposto por Windmeijer (2018) torna viável esse importante diagnóstico. Destaca-se, por fim, que os estimadores aqui avaliados

podem ser adaptados com o intuito de produzir inferências adequadas mesmo na presença de certos problemas de especificação, a exemplo de autocorrelação nos erros do modelo. Nesse sentido, alguns procedimentos de correção de viés são propostos e discutidos por Zhou et al. (2014).

Notas

- 1 É importante observar que o jargão utilizado por diferentes autores para descrever esses métodos de análise pode variar sensivelmente. A terminologia aqui utilizada segue a adotada por Arellano (2003) e, em parte, Wooldridge (2010).
- 2 Outra diferença é que os métodos tradicionais de EF não permitem a inclusão entre os regressores de variáveis que não apresentam variação temporal, ao contrário dos procedimentos de EA.
- 3 A validade de certas suposições de exogeneidade sequencial dependerá do padrão de autocorrelação (ou da ausência de autocorrelação) dos erros u_{it} , o que pode ser testado formalmente pelo pesquisador.
- 4 Essa modelagem também assume implicitamente que o pesquisador incluiria quaisquer variáveis w_{1it}, \dots, w_{kit} disponíveis (isto é, observadas) e correlacionadas simultaneamente com y e com x , conforme discutido na seção 2.1.1.

Referências

- Angrist, J. D., & Pischke, J. (2008). *Mostly harmless econometrics: An empiricist's companion*. New Jersey: Princeton University Press.
- Arellano, M. (2003). *Panel Data Econometrics*. New York: Oxford.
- Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(194), 277–297.
- Bertrand, M. and Schoar, A. (2003). Managing with style: The effect of managers on firm policies. *Quarterly Journal of Economics*, 118(4), 1169–1208.
- Black, B., & Kim, W. (2012). The effect of board structure on firm value: A multiple identification strategies approach using Korean data. *Journal of Financial Economics*, 104(1), 203–226.

- Blundell, R. and Bond, S. R. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115–143.
- Blundell, R. and Bond, S. R. (2000). GMM estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19(3), 321–340.
- Bond, S. R. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal*, 1(2), 141–162.
- Bun, M. J., & Sarafidis, V. (2015). Dynamic panel data models. In Baltagi, B. H., editor *The Oxford handbook of panel data*, 76–110. Oxford: Oxford University Press.
- Dang, V. A., Kim, M., & Shin, Y. (2015). In search of robust methods for dynamic panel data models in empirical corporate finance. *Journal of Banking & Finance*, 53, 84–98.
- Fama, E. F. and French, K. R. (2002). Testing trade-off and pecking order predictions about dividends and debt. *Review of Financial Studies*, 15(1), 1–33.
- Flannery, M. J., & Hankins, K. W. (2013). Estimating dynamic panel models in corporate finance. *Journal of Corporate Finance*, 19, 1–19.
- Frank, M. Z. and Goyal, V. K. (2003). Testing the pecking order theory of capital structure. *Journal of Financial Economics*, 67(2), 217–248.
- Glen, J., Lee, K., and Singh, A. (2001). Persistence of profitability and competition in emerging markets. *Economics Letters*, 72(2), 247–253.
- Greene, W. H. (2000). *Econometric analysis*. Prentice Hall, Upper Saddle River, New Jersey, 4th edition.
- Himmelberg, C. P., Hubbard, R. G., and Palia, D. (1999). Understanding the determinants of managerial ownership and the link between ownership and performance. *Journal of Financial Economics*, 53(3), 353–384.
- Larcker, D. F., & Rusticus, T. O. (2010). On the use of instrumental variables in accounting research. *Journal of Accounting and Economics*, 49(3), 186–205.
- McConnell, J. J. and Servaes, H. (1995). Equity ownership and the two faces of debt. *Journal of Financial Economics*, 39(1), 131–157.
- Roberts, M. R., & Whited, T. M. (2013). Endogeneity in empirical corporate finance. In Constantinides, G. M., Harris, M., & Stulz, R. M. (Eds.). *Handbook of the Economics of Finance*, 493–572. New York: Elsevier.
- Stulz, R. M. (1990). Managerial discretion and optimal financing policies. *Journal of Financial Economics*, 26(1), 3–27.
- Waring, G. F. (1996). Industry differences in the persistence of firm-specific returns. *The American Economic Review*, 86(5), 1253–1265.
- Windmeijer, F. (2018). *Testing over-and underidentification in linear models, with applications to dynamic panel data and asset-pricing models* (Working Paper No. 18/696). Department of Economics, University of Bristol, UK.
- Wintoki, M. B., Linck, J. S., & Netter, J. M. (2012). Endogeneity and the dynamics of internal corporate governance. *Journal of Financial Economics*, 105(3), 581–606.
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). Cambridge, Massachusetts: MIT Press.
- Zhou, Q., Faff, R., & Alpert, K. (2014). Bias correction in the estimation of dynamic panel models in corporate finance. *Journal of Corporate Finance*, 25, 494–513.

Agências de fomento: O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Projeto nº 401562/2006-9.

Autores

1. Lucas A. B. C. Barros, Pós-Doutor em Finanças pela Universidade da Pensilvânia, Philadelphia, USA.

E-mail: lucasbarros@usp.br

ORCID

 0000-0002-9319-1306

2. F. Henrique Castro, Pós-Doutor em Finanças pela Universidade Columbia, Nova York, USA.

E-mail: castrohfg@gmail.com

ORCID

 0000-0001-7456-2354

3. Alexandre Di Miceli da Silveira, Pós-Doutor em Gestão de Negócios pela Universidade Louvain-la-Neuve e pela Universidade Cornell, Louvain, Bélgica. E-mail: alexandre.miceli@fecap.br

ORCID

 0000-0002-4864-9429

4. Daniel Reed Bergmann, Doutor em Finanças pela Universidade de São Paulo, São Paulo, Brasil.

E-mail: danielrb@usp.br

ORCID

 0000-0003-1567-0229

Contribuição dos autores

Contribuição	Lucas Barros	F. Henrique Castro	Alexandre Di Miceli	Daniel Reed Bergman
1. Definição do problema de pesquisa	√	√	√	√
2. Desenvolvimento das hipóteses ou questões de pesquisa (trabalhos empíricos)	√	√	√	√
3. Desenvolvimento das proposições teóricas (ensaios teóricos)				
4. Fundamentação teórica/Revisão de Literatura	√	√	√	√
5. Definição dos procedimentos metodológicos	√	√	√	√
6. Coleta de Dados				
7. Análise Estatística	√	√	√	√
8. Análise e interpretação dos dados	√	√	√	√
9. Revisão crítica do manuscrito	√	√	√	√
10. Redação do manuscrito	√	√	√	√
11. Outra (favor especificar)				