Forthcoming in the third edition of the *Handdbook of Insurance*, 2024

Causality in empirical analyses with emphasis on asymmetric information and risk management†

Georges Dionne
Canada Research Chair in risk management and Finance department
HEC Montréal

2 November 2023

Abstract

We discuss the difficult question of measuring causality effects in empirical analyses, with applications to asymmetric information and risk management. It is now well documented in the economic literature that policy analysis must be causal. Hence, the measurement of its effects must also be causal. After having presented the main frameworks for causality analysis, including instrumental variable, difference-in-differences, and generalized method of moments, we analyze the following questions: Does risk management affect firm value and risk? Do we face a moral hazard problem in the insurance data? How can we separate moral hazard from adverse selection and asymmetric learning? Is liquidity creation a causal factor for reinsurance demand? We show that residual information problems are often present in different markets, while risk management may increase firm value when appropriate methodologies are applied.

***Keywords*:** Asymmetric information, moral hazard, adverse selection, risk learning, risk management, causality test, dynamic data, essential heterogeneity, difference-in-differences, instrumental variable, propensity scor, generalized method of moments.

***JEL numbers*:** C12, C18, C23, C25, C26, D80, G11, G22.

† *The author would like to thank SSHRC-Canada for financial support. He would also like to acknowledge the researchers who helped him develop several of the ideas on the subject over the years: P.A. Chiappori, D. Desjardins, A.T. Fenou, R. Gagné, N. Koné, Y. Liu, P.C. Michaud, M. Mnasri, J. Pinquet, B. Salanié, and C. Vanasse. He thanks Claire Boisvert and Karen Sherman for their collaboration in the preparation of the chapter.*

1 Introduction

This chapter presents the basis of applied econometric analysis for causal analysis and discusses examples of applications. Our goal is to improve the understanding of measurement of economic policy analysis in risk management and insurance. According to Heckman and Pinto (2022), policy analysis must be causal analysis. Causal analysis estimates the factors that generate outcomes, and isolate the role of interventions or external shocks on resource allocation. Causal analysis must control for possible counterfactual problems from the control group with appropriate methodologies in order to isolate the appropriate causal factors.

The applications in this new chapter concentrate on causality methodologies for testing the presence or absence of a residual information problem in different markets with an emphasis on insurance, and for verifying the effect of risk management on firm value and risk. The revised contributions highlight various difficulties that are not always well understood by those who perform the empirical measurement of causal relationships. In fact, conclusions are often made based on correlations instead of on causality.

One application is measuring for the presence of asymmetric information in an insurer's portfolio. Is risk classification sufficient to rule out residual asymmetric information or do we need self-selection mechanisms inside risk classes? We treat the separation issue between moral hazard and adverse selection and explore how dynamic data can be used to develop tests for achieving such separation. We also analyze the effect of a new insurance pricing scheme in the presence of asymmetric information with a DID model and show that moral hazard is not rejected. Other resource allocation applications, such as reinsurance demand and liquidity creation in the insurance industry, CDS central clearing to reduce information problems in OTC markets, and premium in M&A transactions in the presence of asymmetric information are presented with causality analysis.

We will also review the recent literature on the empirical effect of risk management on firm value when essential heterogeneity is potentially present. Does risk management increase firm value or do high value firms engage in more risk management? We shall also look at the relationship between risky firms and risk management.

We do not consider dynamic treatment analysis in this chapter since the covered applications do not contain events and data with dynamic treatments. In fact, dynamic treatment effects have become popular only recently. Basic causality models in applied econometrics do not use dynamic models. For a dynamic treatment effect analysis with instrumental variables or matching modeling see Heckman et al. (2016), and for an analysis of dynamic heterogeneous treatments effects that generalizes the standard DID model see Sun and Abraham (2021).

The rest of the chapter is organized as follows. The remainder of the introduction covers the literature review of asymmetric information in insurance markets and of risk management value. Section 2 presents the fundamental contributions on causality analysis made by three Nobel prize winners in 2021. Section 3 discusses the main issues related to causality in applied econometrics with an emphasis on instrumental variables and essential heterogeneity. Section 4 presents three applications: two with instrumental variables including one with potential essential heterogeneity and one with the use of propensity score matching methodology in a DID application. Sections 5 and 6 are devoted to different tests of asymmetric information in insurance markets using dynamic data and DID analysis. Section 7 presents an application of the generalized method moments to analyze the causal relationship between insurance demand and liquidity creation in a dynamic environment with the GMM, and section 8 concludes the chapter.

1.1 Asymmetric information in insurance markets

The study of information problems in economics began in the early 1960s. The two best known problems, moral hazard and adverse selection, were introduced in the literature in 1963 by Kenneth Arrow in a classic article published in the *American Economic Review*. In 1970, Akerlof proposed the first analysis of market equilibrium in the presence of adverse selection. Optimal contracts were characterized endogenously (security design) for adverse selection in articles by Rothschild and Stiglitz (1976) and Wilson (1977), and for ex ante moral hazard by Holmström (1979) and Shavell (1979). Ex post moral hazard was defined by Pauly (1968) and was later formalized by Townsend (1979) and Gale and Hellwig (1985).

In the early 1980s, several theoretical developments were advanced to account for different facts observed in markets. Specifically, multi-period contractual relations were introduced; the renegotiation of contracts was formalized; the problem of contractual commitments was analyzed; and simultaneous treatment of several information problems became a consideration. Other noteworthy proposals were developed to explain hierarchical relations in firms and in organizations, often involving multi-party participants and contracts in insurance, education, healthcare, and risk management.

Economic relationships in banking and insurance contracts, labor and sharecropping contracts, and auctions were studied. Several forms of contracts observed in these markets were catalogued in various theoretical contributions. The best known are partial insurance coverage (co-insurance and deductibles), compensation based on hours worked and performance, executive compensation with stock options, debt with collateral, bonus-malus schemes, temporal deductibles, and venture capital contracts with warrants. In addition, several corporate organizational practices were rationalized, such as the use of foremen, internal and external controls, auditing, decentralization of certain decisions, and the centralization of more difficult-to-control decisions.

The empirical study of information problems began much later. The main motivation was to distinguish the stylized (qualitative) facts used to construct certain theoretical models from real or more quantitative facts. For example, in theoretical contributions, different automobile insurance deductibles may well be used to reduce adverse selection, but there is no evidence that insurers established this partial coverage for that reason. It can also be argued that labor contracts with performance compensation are used to reduce moral hazard in firms, but it has not necessarily been conclusively empirically demonstrated that there is less moral hazard in firms with this form of compensation than in other firms that use fixed compensation, combined with other incentives or control mechanisms to deal with this information problem.

Another strong motivation for empirically verifying the causal effects of information problems is the search for ways to reduce their negative impact on resource allocation. For example, we know that partial insurance is effective in reducing ex ante moral hazard, as it exposes the insured person to risk. Yet this mechanism may be ineffective against ex post moral hazard, because the accident has already occurred. Partial insurance may even have pernicious effects and encourage the padding of costs (Dionne and Gagné, 2001). The fact that the audit of files seems to be the most effective instrument against ex post moral hazard shows the importance of identifying the real problem when attempting to correct imperfections and improve resource allocation.

When it comes to empirically measuring information problems and assessing the effectiveness of mechanisms set up to correct them (relationship between the nature of contracts and their performance), numerous complications soon arise. For one, several information problems may exist simultaneously in the database studied; the theoretical predictions must then be carefully defined to distinguish the effects of different information problems on the parameters of the models to be estimated. Moreover, firms have a wide range of mechanisms (substitutes or complementary) at their disposal, which may be selected for reasons other than information problems or for information problems other than those investigated in a particular study. In other words, the information problems under consideration are often neither a necessary nor a sufficient condition to justify the existence of certain observed mechanisms.

Treating several information problems simultaneously in empirical studies is difficult: The literature does not yet offer strong theoretical predictions. If we simply examine whether a market contains any residual information asymmetry, regardless of its origin, it is easier to demonstrate its absence because there is no need to distinguish between the different forms of information asymmetry. Otherwise, we have to ascertain which residual form is still present and document its cause to analyze the instruments that could mitigate or eliminate it.

As a rule, the distinction between moral hazard and adverse selection can be reduced to a problem of causality (Chiappori, 1994, 2000). With moral hazard, the non-observable actions of individuals that affect accidents are consequences of the forms of contracts. For example, a non-optimal contract may increase the risk of driving because it reduces the incentives to act safely. With pure adverse selection, the nature of different risks already exists before contracts are written. The contracts selected are dictated by the risks present. There is thus a potential form of reverse causality between the two information problems. When an exogenous change occurs in an insurance market, we can limit our test to the way it affects existing policy holders in the treatment group and isolate a moral hazard effect if insured in the treatment and control groups are almost identical. Alternatively, we could make comparisons to see whether the chance of accident differs between new and old policy holders and check for any bias caused by adverse selection. Another approach is to use dynamic models and develop causality tests. However, these tests must consider that other information asymmetries may be present such as the learning of the contract parties over time (Dionne et al., 2013a, 2013b). Learning may or may not be symmetrical. Dynamic data are also useful for separating moral hazard from unobserved heterogeneity (Abbring et al., 2003; Dionne et al., 2011). Even if standard causality models in the literature (Arkhangelsky et al., 2021; Heckman et al., 2006) use panel data, they do not necessarily incorporate dynamic behavior in their analysis.

Another difficulty in the empirical measurement of information problems is the fact that researchers are not privy to more information than are decision makers. On the contrary, they often have less information. Two solutions have been adopted to mitigate that difficulty: (1) use of confidential surveys and (2) development of econometric strategies that can isolate the desired effect. The experimental approach is a third avenue that I shall not deal with in detail.

The survey method has the advantage of providing direct access to private information not available to one party to the contract, such as accidents not claimed or risk perception. Such information makes it possible to measure motivations for choosing specific contractual clauses directly, along with agents' behavior. The drawback of this method is that it is very costly. It can also be biased, because it is very difficult to explain the complexity of the problem studied to participants, and because several alternative explanations, such as risk aversion, might have been overlooked in the questionnaires. Another source of bias is related to the formation of representative samples.

The development of econometric strategies requires knowledge of the theoretical problem under study and of the econometric methods suitable for the project. This is why the most productive research teams are composed of theoreticians and econometricians. The objective is to isolate effects that are not directly observable by both parties to the contract but that are taken into account by certain variables or combinations of variables. As discussed by Chiappori (1994, 2000) and Gouriéroux (1999), econometric research consists in distinguishing between two sources of information. The first type is composed of variables observable by the two parties to the contract. These variables can be used to make estimates conditional on the characteristics observed. The second type is linked to the information that is not observable by econometricians (and by at least one contractual party), but that may explain choices of contracts or behaviors. In the case of adverse selection, choices of contracts can be interpreted by econometricians as being an endogenous selection. One way to take this into account is to estimate agents' decisions simultaneously by introducing hidden connections (or informational asymmetries) between the decisions. One known test is the non-null correlation between the random terms of the different equations (contract choice and accident distributions; Chiappori and Salanié, 2000). Another test entails estimating the parameters of contract choice on the contract result (Dionne et al., 2001). However, these early studies do not identify the causality effect of information problems on resource allocation (Dionne et al., 2009). They are limited to the presence or absence of residual asymmetric information and to correlations of contract forms and accidents. They cannot conclude that accidents cause contract forms (adverse selection) or that contract forms cause accidents (moral hazard).

Quality of data is a determining factor in the measurement of desired effects. The data must correspond directly to the contractual relations studied and to the duration of the contractual periods. There must also be access to data broken down contract by contract. The effort involved in formulating raw data for research purposes should not be underestimated. Raw data are used in the day-to-day operations of firms that are not concerned with research problems, and do not always contain direct information on variables needed for the problem studied.

Econometric specifications must correspond to the theoretical models under consideration, if erroneous conclusions are to be avoided. Often, researchers choose (or are forced) to use only part of the information available to decision-makers, and thus bias the effects of certain treatments such that they capture the effects of other forgotten or inaccessible variables (unobserved counterfactuals) and obtain false conclusions.

Finally, the participants to different contracts are often risk averse to varying degrees. This characteristic is also difficult to observe and can be a source of asymmetric information. Some authors have recently proposed models that take into account the varying degrees of risk aversion (Cohen and Einab, 2007), but very few predictions can isolate the effects of information problems as they relate to varying degrees of risk aversion among agents.

1.2 Risk management value

When there are no market imperfections, market prices contain all information, making it impossible to generate a profit based on informational advantages. Although this result is widespread, many managers continue to believe that they possess comparative advantages in certain markets. Consequently, firms use their resources to develop investment strategies that are risky because a high return is generally accompanied by a high risk. However, these practices are not followed by firms that realize they do not actually possess comparative advantages within their sector or those that had bad experiences resulting from the inappropriate use of hedging instruments. In fact, firms do not necessarily need to hedge against all the financial risks they may face, particularly when they are already well diversified internally.

The main goal of risk management is to increase firm value by reducing the cost of risk when there are market imperfections. The four main sources of market imperfections are default costs, agency costs, investment financing, and taxes. Dividend payment, managers' risk attitude and corporate governance problems may also explain risk management of non-regulated firms (Dionne and Ouederni, 2011; Dionne, 2019).

Market imperfections generate default costs. Default costs refer to the costs associated with default, not bankruptcy (Stulz, 1996). Default costs can be divided into two categories: direct costs such as lawyer fees, consultant fees and court-related expenses, and indirect costs incurred when a firm is under bankruptcy protection laws, such as reorganizational costs. These two categories of costs are directly reflected in a firm's valuation. The goal of an efficient risk management strategy is to maintain these costs at an optimal level, while taking into consideration the cost of hedging activities.

Risk management can allow a firm to reduce the expected tax payments when the taxation function is convex with respect to profits or firm value (Graham and Smith, 1999; Graham and Rogers, 2002). A good risk management strategy may increase a firm's debt capacity and its capital structure. In other words, risk management can be interpreted as a substitute for equity, by reducing the default probability and hence the default risk premium imposed by banks or investors. By reducing the risk premium, hedging can create new investment opportunities financed by debt (Smith and Stulz, 1985; Campello et al, 2011; Dionne and Triki, 2013). Inversely, capital structure can also impact how a firm approaches risk management (Stulz, 1996); as a result, causality analysis is essential to establish the real effects.

Under asymmetric information, external financial costs of investment are much higher than internal financial costs (Froot et al, 1993). This situation increases the incentives to protect internal financing with risk management. Firms whose managers are also shareholders (meaning that they also benefit from the firm's profits) are usually poorly diversified. Tufano (1996) tested this premise for firms in the gold mining industry. He found that managers who have a large portion of their human capital and compensation invested within their firm seek to protect themselves more by using firm risk management. Attributing firm equity to managers is beneficial when it comes to risk management, yet this incentive can often be more costly than stock options. Stulz (1996) explains why firms that compensate managers with stock options may be more lax with respect to risk management since risk management may reduce the probability of the option being in the money at the end of the period and the possibility of making money with options. Good corporate governance reduces the agency costs of risk management between the board and the managers (Dionne et al., 2019).

There are many other motivations for firm risk management. They include lack of liquidity, mergers and acquisitions, higher productivity in producing goods and services, and other strategic behaviors (Dionne, 2019). The main question is the extent that risk management increases firm value and reduces its risk. We will see below that the current empirical evidence is ambiguous. We argue that this can be due to methodological problems such as the absence of causality analysis.

2 Three Nobel lectures in economics on causality

2.1 Basic model

The three 2021 Nobel lectures in economics were on causality analysis. Establishing causal effects is very important for decision makers. Three distinct literatures proposed developments of methodologies for estimating causal effects: those in statistics, econometrics, and computer science.

Research on causality in statistics started in the first half of the twentieth century (Imbens, 2022, Nobel lecture). It was not until the 1990s that the term causality began to become popular in applied econometrics and statistics. Before that period, correlation was often associated with implicit causality without appropriate methodology.

The research in the statistical literature started with randomized experiments where the treatment and the control groups were formed randomly by applying, for example, the Randomized Controlled Trials (RCTs) methodology. Randomization simplifies the comparison of treatment and control groups before the treatment. Randomization balances the covariates (observable and non-observable) between the two groups. However, many real life applications do not correspond to randomization. Empirical studies with different groups are more common.

Let us start with an example presented by Imbens (2022), where causality is not limited to testing and estimating causal effects in randomized experiments. Rubin (1974) defined the causal inference problem where one compares two potential outcomes defined for the same agent *i*. $Y_i(C)$ is the agent's outcome in the control group, and $Y_i(E)$ is the outcome for the same agent in the experimental group, at the same time, with exposure to the treatment. The causal effect is the difference $Y_i(E)-Y_i(C)$, or the comparison of outcomes from a treatment and a control situation. The average causal effect over the population is equal to:

$$\frac{1}{N}\sum_{i=1}^{N}\left(Y_i(E)-Y_i(C)\right), \tag{1}$$

where N is the number of agents in the studied population. This causal example has an identification problem: the two potential outcomes cannot be observed for the same agent at the same time. Holland (1986) referred to this as the fundamental problem of causal inference, implying that causal analysis must be performed on different groups or populations, which complicates the exercise because the two groups must be comparable in order to isolate a causal effect. The composition of the groups should not be affected by the treatment, and non-observable characteristics of the subjects should be balanced between the groups. In a study with randomized group formation, the two groups obtained from a fair lottery on a given population are similar before the treatment if the lottery is well managed.

Rosenbaum and Rubin (1983a, 1983b) focused on the case where assignment to treatment is not completely random, but conditional on some observed confounders (covariates) that may affect both the treatment and the outcome. If the observed confounders for agent *i* are denoted by X_i, the absence of confoundedness is measured by the conditional independence on X_i of the treatment W_i and potential outcomes Y_i. Let $W_i \in \{C, E\}$, we can write:

$$W_i \perp \left(Y_i(C), Y_i(E)\right) | X_i. \tag{2}$$

where $\perp$ is for independence. The distinction between confounders or covariates X_i and causal variables W_i is important in this framework. The absence of confoundedness is an assumption (to be tested) in the process that determines the effect of W_i on outcomes Y_i. Without this absence, one cannot draw any conclusions about the causality of the treatment.

Formally, unconfoundedness requires that the probability of treatment assignment is free of dependence on the potential outcomes. "Outcome under the control treatment, $Y_i(C)$, given active treatment and given covariates, is identical to its distribution conditional on control treatment and conditional on covariates, and second, that, analogously, the conditional distribution of the outcome under the active treatment $Y_i(E)$, given receipt of the control treatment and conditional on covariates, is identical to its distribution given the active treatment and conditional on covariates" (Imbens and Rubin, 2015). Informally, unconfoundedness requires a sufficiently rich set of pre-treatment variables that remove systematic biases from comparisons between treatment and control units. The unconfoundedness assumption is not directly testable. The issue is that the data are not directly informative about the distribution of the control income $Y_i(C)$ for those who received the active treatment (for those with $W_i = E$). Further, the data is not directly informative about the distribution of the active treatment outcome given receipt of the control treatment (for those with $W_i = C$, for which we never observe $Y_i(E)$). Thus, the data cannot directly provide evidence to support the validity of the unconfoundedness assumption. A non-observed covariate or confounder could explain the treatment result. Causal estimations can be used to isolate the effect of the treatment. The instrumental variable approach can be useful under assumptions more restrictive than assuming unconfoundedness. Replication is another way to validate the treatment result when the first study used a small sample. It is a method that reduces bias in the original study.

This statistical approach differs from the early studies in econometrics where treatments stem from economic agents' welfare maximization behavior. Leamer (1983) and Lalonde (1986) influenced the development of causality in applied econometric studies significantly, such as the research of the Princeton labor group (Ashenfelter, Card and

Krueger) and the credibility revolution of causality analysis in the econometric literature, to which Imbens and Angrist have contributed.[1]

2.2 Local Average Treatment Effects (LATE)[2]

By testing which instrumental variable (IV) estimates are valid, the local average treatment effects (LATE) framework is appropriate to measure causality under certain circumstances. In his Nobel lecture, Angrist (2022) presented empirical examples to illustrate the value of the LATE framework for causal inference. LATE identifies independence conditions satisfied by random or non-random assignment from other exclusion restrictions.

An empirical strategy for policy evaluation is a research strategy that includes data collection, identification of causal factors, and estimation. Angrist (2022) considers the Randomized Clinical Trials (RCTs) model a powerful research approach. As mentioned above, random assignment ensures that treatment and control groups are comparable in the absence of treatment. As a result, post-treatment differences in average outcomes reflect the treatment effect.

Angrist (2022) identified instrumental variable (IV) methods and regression discontinuity (RD) designs for RCTs when experimental random assignment is not possible. In applications of IV and RD, causal variables of interest are referred to as treatment variables.

Angrist (2022) presented a contribution by Bloom (1984) that inspired the LATE model. Consider a clinical trial that offers a treatment randomly. Proportion π receives the treatment, while the rest of the group does not. Those who are offered treatment are identified with the dummy variable Z_i, and those who receive the treatment are identified with the dummy variable D_i. Potential outcomes for subject i in the treated and untreated groups are denoted by Y_{1i} and Y_{0i}, respectively. The observed outcome is:

[1] Well-known examples of early studies include those of Angrist (1990), Angrist and Krueger (1991), Card (1990), Ashenfelter and Krueger (1994), Card and Krueger (1994), and Imbens and Angrist (1994).

[2] LATE is a discrete approximation of the Marginal Treatment Effect (MTE) discussed below (Heckman et al., 2006).

$$Y_i = Y_{0i} + D_i\left[Y_{1i} - Y_{0i}\right]. \tag{3}$$

$Y_{1i} - Y_{0i}$ measures the causal effect of treatment on individual *i*. Bloom (1984) shows how to compute the average effect on the treated in this scenario. Let *δ* be the causal effect of treatment assigned on Y_i. Then,

$$E\left[Y_{1j} - Y_{01} \middle| D_i = 1\right] = \frac{\delta}{\pi}. \tag{4}$$

This is an IV experiment that uses treatment assigned as an instrument for treatment received $\left(D_i\right)$.The LATE model generalizes this result.

Imbens and Angrist (1994) were among the first to propose conditions for considering essential heterogeneity, and developed the Local Average Treatment Effect (LATE) measure of policy intervention. The local average treatment effect can be measured as:

$$\text{LATE} = E\left(Y_i\left(E\right) - Y_i\left(C\right) \middle| W > 0\right) \tag{5}$$

where *W* is a policy variable. The LATE theorem generalizes Bloom's result by adding the monotonicity condition.

One concern is that LATE estimates the average effect for a subpopulation that may not be identified. Moreover, LATE is not necessarily invariant to the assignment mechanism. Some inferences can be made, however, about this subpopulation (Abadie, 2003).

Angrist and Imbens (1995) extended the basic model to multiple endogenous variables. Motivated by Angrist and Krueger (1991), Angrist and Imbens (1995) also extended the LATE model to ordered treatments such as years of schooling, while Angrist et al. (2000) added continuous treatments and simultaneous equations models.[3]

[3] These extensions are discussed in Imbens (2022).

2.3 Instrumental variable may not be sufficient

To better identify the average treatment for a subpopulation, Angrist (1990) proposed the following instrumental variable application. The author was interested in the causal effect of serving in the military ($W_i = 1$) on earnings (Y_i). The simple comparison of earnings between veterans $Y_i(1)$ and non-veterans $Y_i(0)$ may not be sufficient to establish a causality effect of serving in the military. Omitted covariates may cause bias in a comparison of veterans and non-veterans' earnings, even when controlling for various observed characteristics of the individuals under the assumption of the absence of confounders.

Angrist (1990) used draft eligibility as an econometric instrument (Z). It is clear that the instrument can affect the probability of serving in the military. If the instrument is correlated with the endogenous variable of serving in the military and not with estimated earnings, it satisfies the relevancy condition (Staiger and Stock, 1997). It could have also satisfied the exclusion restriction, namely that the only effect of being draft-eligible was serving in the military. Even if, from standard econometric modeling, this approach seems adequate, it is not clear that it is sufficient to isolate the causal effect of military service on earnings.

Indeed, Heckman (1990) and Manski (1990) showed that it is not possible to identify the average causal effect of military service without additional assumptions. Heckman (1990) argued that identification of this average effect requires z values for the instrument, such that the probability of being a veteran, conditional on the instrument, $P(W_i = 1 | Z_i = z)$ is arbitrarily close to zero or one. Manski (1990) defined sample bonds with the same properties as the condition proposed by Heckman (1990). These contributions motivated the development of the monotonicity condition by Imbens and Angrist (1994).

The development of methods to isolate causal effects have inspired research in statistics on instrumental variables and research in econometrics on matching methods and experimental design. These developments are well described in the works of Imbens (2022) and Angrist (2022).

2.4 Propensity score

The rise of applied econometrics generated a wave of theoretical econometric innovations. One methodology application built on Rosenbaum and Rubin's (1983b) propensity score theorem. Propensity score is the probability that an individual with different covariates is assigned to the treatment group. This theorem contributed to the credibility revolution of causality analysis in the econometric profession by shifting econometricans' attention to the foundations of treatment analysis rather than simply estimating models for outcomes. Dehejia and Wahba (1999) were the first to demonstrate the value of the propensity score for applied work, while Hirano et al. (2003) and Angrist and Hahn (2004) addressed important theoretical questions about the score model. We present an application of the propensity score approach in Section 4.

2.5 Difference-in-differences (DID)

Transparency around the critical assumptions is crucial when studying causality in economics. Understanding the limits of what data can credibly tell us and making the research accessible to a broader social science community were important goals for David Card. In his Nobel lecture, Card (2022) presented the design-based perspective approach, with an emphasis on understanding the assignment mechanism. He contended that design-based studies put causality analysis front and center in research projects. Design-based studies are particularly useful for testing basic predictions of a theory or testing between competing theories.

For Card (2022), design-based research is a solution to the problems of credible inference identified by Ashenfelter and Heckman (1974), Leamer (1978), and Hendry (1980), among others. Ashenfelter and Card proposed the term difference-in-differences in 1985, in a paper in which they identified problems with the newest generation of models that were supposed to separate causal effects from selection biases in longitudinal models. Does the control group evolve in the same way as the treatment group before the treatment date? The parallel trends test can verify this. Card contributed to the definition and importance of control groups. An application of a DID analysis with parallel trends test is presented in Section 6.3.

In most design-based studies, as in a classical Randomized Controlled Trials study (RCTs), the counterfactual state for treatment is the state experienced by the comparison group. In RCT studies, the groups are often randomly chosen. One concern, for Card (2022), is that the particular counterfactual identified by the design may be too restrictive.

Card was also very active in discussions with practitioners (Goux and Maurin, 2023). He contributed to many debates on subjects such as the minimum wage, unemployment insurance effects, and inequalities in wages in developed countries. Education and immigration were important research topics for him.

2.6 Synthetic Control (SC)

Synthetic Control (SC) is intended to improve the quality of the control or comparison group when the data do not permit robust parallel trends analysis (Abadie and Gardeazabal, 2003; Abadie 2021; Abadie and L'Hour, 2021; Abadie et al., 2010, 2015). The goal is to obtain the best control group that reproduces the counterfactual outcome that the treated agents would have experienced in the absence of the treatment. The idea behind the SC method is to use information from many potential control groups instead of only one to ensure a better comparison with the treated group. The created SC group is a weighted average of chosen control groups.

2.7 Synthetic difference-in-differences (SDID)

A new estimator for causal effects with panel data that builds on difference-in-differences (DID) and synthetic control (SC) methods was recently published (Arkhangelsky et al., 2021). The authors found, theoretically and empirically, that the synthetic difference-in-differences (SDID) estimator has robustness properties, and that it performs well where other estimators are commonly used. They studied the asymptotic behavior of the estimator when the systematic part of the outcome model includes latent unit factors that interact with latent time factors.

As Currie et al. (2020) demonstrated, DID methods have been widely used in applied economics over the last three decades. More recently, SC methods have emerged as an important alternative method for comparative case studies. DID methods are often applied in cases with a large number of agents or units that are exposed to the policy in a panel data study. The parallel trends assumption must be tested in this environment. SC methods, often used in applications with small numbers of agents exposed, compensates for the lack of parallel trends by weighting control groups. Similar to SC, the SDID method weights and matches pre-exposure trends of different control groups. Like DID, the SDID method allows for valid large-panel inference. The model also extends the difference-in-differences approach by permitting the effect of unobserved factors on outcome to vary with time. Dionne et al. (2023) presents an application of the model.

2.8 Regression Discontinuity (RD)

Regression Discontinuity was introduced by Thistlethwaite and Campbell (1960) for estimating treatment effects in a nonexperimental setting, and was used intensively by Angrist (2022) in different studies of attendance at school. The treatment is obtained when an observed assignment variable D exceeds a known cutoff point c. Consider the example in which Thistlethwaite and Campbell (1960) analyzed the impact of merit awards on future academic outcomes (Lee and Lemieux, 2010). The treatment was given to agents with a score value D greater than or equal to c. Individuals with a score value just below the cutoff are quite comparable to those just above the cutoff, so they can be considered to form the control group.

It is clear that agents should not be able to manipulate D, the causal variable. With no manipulation, the variation of the treatment near c can be considered random, as in a randomized experiment. Other covariates do not matter in that case, contrary to the IV and DID regressions, where tests must be done on potential counterfactuals to validate the causal conclusions. In that sense, RD is a less restrictive research design, which may explain why it became very popular in recent years.

2.9 Causality and computer science

In terms of causality within the computer science literature, Pearl (2009) developed Directed Acyclic Graphs (DAGs) and Structural Causal Models (SCMs) to study causality.

According to Imbens (2022), the DAG approach has two distinct benefits. The first is pedagogical, by formulating the critical assumptions that capture the causal relationships. DAG can be a powerful way to illustrate the key assumptions underlying causal models. A second potential benefit lies in the mathematical tools developed in the recent DAG literature. For example, Do-calculus, developed by Pearl (2012), can be used to answer causal questions differently than in econometrics.

3 Causality in applied econometrics

3.1 Introduction

An econometric model is based on interpretable behavior that come from economic theory, where economic agents maximize their behavior. The model allows one to test the sources of potential outcomes. One of its activities consists of constructing credible policy counterfactuals, including the forecasting of policy impacts in new environments. The econometric approach considers the definition of causal parameters and their identification as two different tasks.

Two well-known approaches (the Neyman-Rubin-Holland approach in statistics and the Do-calculus approach in computer science) address some of the same problems as those in the econometric approach. All approaches start from the basic intuitive definition of a causal effect as a ceteris paribus consequence of a policy change. However, the rules of constructing and identifying counterfactuals in statistics and computer science are very different from those in econometric analysis.

Heckman and Pinto (2022) presented the econometric causal model for policy analysis developed from the seminal contributions of Frisch (1930, 1938) and Haavelmo (1943, 1944). They compared the econometric causal approach with the two other causal frameworks discussed above: the Neyman-Rubin (1923) and Holland (1986) (NH) causal model and the Do-calculus (DoC) causal model (Pearl, 2012).

Heckman and Pinto (2022) argued that economists who use these two causal frameworks often discount the benefits of the econometric causal approach. These two models may have limitations when applied to problems analyzed by economists. The NH approach does not consider unobservable variables and some restrictions on empirical relationships derived from economic theory. In contrast, the DoC approach does not treat the functional restrictions and covariance information used in econometrics. It cannot incorporate monotonicity and the separability restrictions that are important components of the instrumental variable analysis. For example, the Generalized Roy model cannot be identified with this approach.

We now present the causality framework proposed by Heckman and Pinto (2022).

3.2 Causality and mapping

Heckman and Pinto (2022) start with a formal definition of causality that relies on mapping inputs X (usually a vector) to an output Y. A map is stable if changing its arguments over the domain of X preserves the mapping relationship. A simple linear model obtained from a map between X and Y is:

$$Y = \alpha + \beta X. \tag{6}$$

Stability means that α and β are not affected when X is changed. This is the invariance of relationships concept proposed by Frisch (1938). Directionality is central for causality analysis. Inverting a map may produce a stable relationship that is not necessarily causal. In their framework, other potential outcomes associated with different values of X are defined as counterfactuals associated with X. The causal effect is the comparison of Y values for different values of X, ceteris paribus.

3.3 Regression analysis

Equation (7) presents a simple regression of Y on T where Y and T are observed variables and U is an unobserved variable:

$$Y = T\beta + U. \tag{7}$$

In terms of (6), $X=(T,U)$. The random link between T and Y operates through two channels: β and $E(U|T=t)$. The presence of U complicates causal analysis because U is not observable.

Heckman and Pinto (2022) considered four potential causal models, as illustrated in Table 1, where the different ε variables are independent external random variables not caused by T,U,Y.

Table 1: Four Potential Causal Models

Causal Model 1	Causal Model 2	Causal Model 3	Causal Model 4
$T=f_T(\varepsilon_T)$	$T=f_T(\varepsilon_T,\varepsilon_V)$	$T=f_T(\varepsilon_T,U)$	$T=f_T(\varepsilon_T)$
$U=f_U(\varepsilon_U)$	$U=f_U(\varepsilon_U,\varepsilon_V)$	$U=f_U(\varepsilon_U)$	$U=f_U(\varepsilon_U,T)$
$Y=T\beta+U$	$Y=T\beta+U$	$Y=T\beta+U$	$Y=T\beta+U$

Source: Heckman and Pinto (2022).

In Model 1, T does not cause U, nor does U cause T. β measures the causal effect of varying T on Y for a fixed value of U. Since variables T and U are independent, the parameter β can be estimated by OLS. In the second causal model, once again, T does not cause U, and U does not cause T. Parameter β is still the causal effect of T on Y. T and U are no longer statistically independent because they are related to a common confounding variable ε_V. The OLS estimator of β could be biased. In Model 3, U causes T and the causal effect of T on Y remains β. In the fourth model, T causes U and the OLS estimator of the parameter β does not identify this causal effect because T also affects Y. The OLS estimator of β captures both the direct and indirect effects of T on Y.

As observed, the OLS model is not necessarily related to a causal interpretation. The OLS estimates the value for Y conditioning on $T=t$. It evaluates the conditional expectation $E(Y|T=t)$ instead of the causal expectation $E(Y(t)|T=t)$, where $Y(t)$ is the value of Y when T is externally fixed to a value t in a causality analysis.

The econometric approach to causality develops explicit hypothetical models in which inputs cause outcomes. According to Heckman and Pinto (2022), the mechanisms governing the choice of inputs is central to study the causal effect of a treatment on outcomes. Identification of empirical counterparts to the hypothetical counterfactuals require careful analysis of unobserved variables (*U*) that cause both input choice and outcomes. Structural econometric models can do that.

Heckman and Pinto (2022) present, in Table 2, three tasks of econometric causal policy analysis. This framework, inspired by the works of Haavelmo (1943, 1944), distinguishes the types of policy analysis where model creation does not require statistical analysis, identification is based on probability theory and estimation implies statistical analysis.

Table 2: Three Distinct Tasks In Causal Policy Analysis

Task	Description	Requirements	Types of analysis
1: Model Creation	Defining the class of hypotheticals or counterfactuals by thought experiments (models)	A scientific theory: A purely mental activity	Outside statistics; Hypothetical worlds
2: Identification	Identifying causal parameters from a hypothetical population	Mathematical analysis of point or set identification: A purely mental activity	Probability theory
3: Estimation	Estimating parameters from real data	Estimation and testing theory	Statistical analysis

Source: Heckman and Pinto (2022).

3.4 Generalized Roy model

3.4.1 The model

Let us consider the Generalized Roy model as an example of a structural model. The Generalized Roy model of counterfactuals analyzes earnings in two sectors of the economy. All agents have two potential incomes: $Y(0)$ in Sector 0 and $Y(1)$ in Sector 1. Agents choose a sector from their potential benefit $I = Y(1) - Y(0)$. Potential incomes depend on observed variables *X,* while *I* may depend on *X* and on an external variable *Z*, which may be a policy variable that influences participation. The agent's choice of Sector 1 $(T=1)$ is identified by $T=1\left[I(X,Z)>0\right]$.

The Generalized Roy model can be represented by four variables $\{Z,V,T,Y\}$. Z is an external policy variable that causes the treatment *T*, which causes the outcome *Y*. *Z* is an instrumental variable. It causes *Y* only through its effects on *T*. *V* contains confounding variables that jointly cause *T* and *Y*. Only variables *Z*, *T*, *Y* are observed. *V* is a source of bias in treatment choice, which makes the evaluation of the causal effect of *T* on *Y* more difficult. The observed relationship between *T* and *Y* may be due to the common effect of *V* on both *T* and *Y*, instead of the causal effect of *T* on *Y*. The average treatment effect (ATE) is the mean treatment effect over all agents, knowing that each agent cannot be in both sectors at the same time: $ATE=E\big(Y(1)-Y(0)\big)$.

Equations (8) to (11) represent the Generalized Roy model. The four error terms (ε) are independent so $Z\perp V$ and $Y\perp Z\,\big|(T,V)$. Equation (8) states that the identification of causal effects in the Generalized Roy model entails controlling for the unobserved confounding variables *V*. A popular approach for doing so uses instrumental variables (IV) that are independent of *V*. They control for *V* by shifting *T* without affecting the distribution of *V* as in (10).

$$V=f_V(\varepsilon_V), \tag{8}$$

$$Z=f_Z(\varepsilon_Z), \tag{9}$$

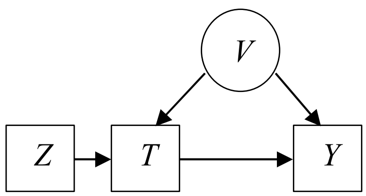
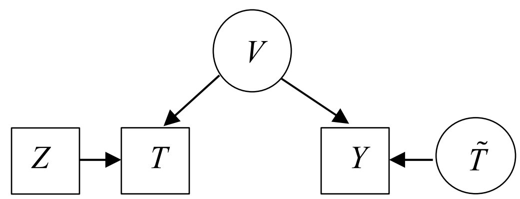
$$T=f_T(Z,V,\varepsilon_T), \tag{10}$$

$$Y=f_Y(T,V,\varepsilon_Y). \tag{11}$$

Table 3, from Heckman and Pinto (2022), represents the empirical Generalized Roy model. Directed arrows identify causal relationships, circles denote unobserved variables, and squares denote observed variables. Hypotheses for each variable of the model are presented in the LMC column.

Table 3 also represents the hypothetical causal model developed by Frisch (1938). This model has the same structure as the empirical Generalized Roy model. It adds an external variable $\tilde{T}$ that replaces *T*. Both variables are independent. $\tilde{T}$ is directly related to *Y* in the hypothetical model. It replaces the link between *T* and *Y*. The hypothetical model is related to experiments while the empirical model is generated from data.

Table 3: Generalized Roy model: Empirical and Hypothetical Causal Models

	Empirical model	Hypothetical model
	 	
	LMC	LMC
V:	$V \perp Z$	$V \perp \left(Z, \tilde{T}\right)$
Z:	$Z \perp V$	$Z \perp \left(V, Y, \tilde{T}\right)$
T:	$T \perp \varnothing \middle| \left(Z, V\right)$	$T \perp \left(\tilde{T}, Y\right) \middle| \left(Z, V\right)$
Y:	$Y \perp Z \middle| \left(T, V\right)$	$Y \perp \left(Z, T\right) \middle| \left(\tilde{T}, V\right)$
$\tilde{T}$:	(not defined for the model)	$\tilde{T} \perp \left(T, V, Z\right)$

$\perp$ means independent of and LMC means Local Markov Condition. The directions and illustrations are from DAGs (Directed Acyclic Graphs).
Source: Heckman and Pinto (2022).

Counterfactuals are obtained by external manipulations of treatments. These are produced in the hypothetical model by conditioning on the variable $\tilde{T}$. For instance, the distribution of the counterfactual outcome *Y*, when the treatment is externally set to a value *t*, is $P_h\left(Y \middle| \tilde{T} = t\right)$ and the counterfactual outcome mean is given by $E_h\left(Y \middle| \tilde{T} = t\right)$. These values are different from those in the empirical counterpart $P_e\left(Y \middle| T = t\right)$ and $E_e\left(Y \middle| T = t\right)$.

Counterfactuals are said to be identified when they can be expressed in terms of the probability distributions of the observed data in the empirical model. This identification requires the econometrician to connect the probability distribution of the hypothetical model with the probability distributions of the empirical model.

3.4.2 Counterfactuals in the Generalized Roy model

The Generalized Roy model can be used to identify counterfactuals. Heckman and Pinto (2022) described several of these approaches. We limit ourselves to the IV model and the Matching model and we present three applications. The standard IV model described by equations (10') and (11') with Z as an instrument cannot identify all counterfactuals without additional assumptions:

$$T = \alpha_0 + \alpha_1 Z + \varepsilon_T \tag{10'}$$

$$Y = \beta_0 + \beta_1 \overline{T} + \varepsilon_Y \tag{11'}$$

where β_1 measures the causal effect of $\overline{T}$ on Y and $\overline{T}$ is the predicted value of T conditional on Z.

The Generalized Roy model is not necessarily captured by this system of two equations often used in applied econometrics. The causal effect, $\beta_1 = Y(1) - Y(0)$ is, in general, a random variable and not a constant, so that treating the causal parameter in equation (11') as a constant does not capture the essential heterogeneity of treatment effects across agents or units. In many applications, the causal effect is stochastically dependent on $\overline{T}$.

3.4.3 LIV model

Heckman and Vytlacil (1999, 2001, 2005) and Heckman et al. (2006) presented the IV problem by introducing a separable choice equation. Their approach enables analysts to control for V and to identify counterfactual outcomes.

The Local Instrumental Variable (LIV) model considers a binary treatment $T \in \{0,\ 1\}$. The separability assumption states that treatment is given by a latent equation that includes the instrumental variable Z and the confounder V. Separability enables us to rewrite the choice equation as:

$$T = 1\left[P(Z) \geq U\right]; \quad P(Z) = P_e\left(T = 1 \middle| Z\right), \tag{12}$$

where $P(Z)$ is the propensity score obtained from the Probit model.

The LIV model (Heckman and Vytlacil, 1999) can be used to identify the distributions of counterfactual outcomes conditioned on U by taking the derivative of the observed outcome with respect to the propensity score. More generally, the counterfactual expectation is identified if there is sufficient variation of the propensity score $P(Z)$ around the value $u \in (0,\ 1)$.

If $P(Z)$ has full support, the average treatment effect (ATE) can be evaluated by:

$$ATE \equiv E_h\left(Y \middle| \tilde{T}=1\right) - E_h\left(Y \middle| \tilde{T}=0\right) = \int_0^1 \left(E_h\left(Y \middle| T=1, U=u\right) - E_h\left(Y \middle| T=0, U=u\right)\right) du. \tag{13}$$

The solution of the LIV model implies a separability assumption that generates a score U for V. Equation (13) assumes that the sample propensity score has enough variation around the value $u \in (0,1)$. The equation is not directly applicable to discrete instruments. Discrete instruments are discussed by Heckman and Pinto (2022).

3.4.4 Matching model

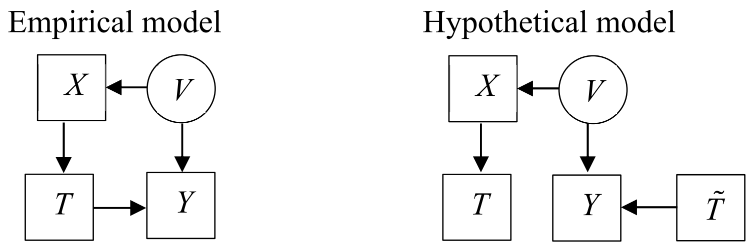
Another method for identifying treatment effects assumes that a set of observed pre-treatment variables suffices to control for the confounding variable V. Matching assumes that the observed variable X is a balancing score for the confounding variable V.

The average causal effect of a binary treatment $T \in \{0,1\}$ is evaluated by the weighted average of mean difference between the treated and non-treated participants that match on X, namely:

$$ATE = \int \left(E_e\left(Y \middle| T=1, X=x\right) - E_e\left(Y \middle| T=0, X=x\right)\right) dF_{e,X}(x) \tag{14}$$

where e is used to identify the matching relationships.

Table 4: Matching model: Empirical and Hypothetical Causal Models



Source: Heckman and Pinto (2022).

The matching assumption replaces the unobserved variable *U* of the Generalized Roy model in Table 4 by the observed variable *X*. In practice, it assumes that potential bias can be eliminated by controlling for observed pre-treatment variables between the agents in the treated and non-treated groups. Under matching, the identification of treatment effects does not require an instrumental variable or separability. This assumption enables us to solve the problem of selection bias induced by unobserved variable *V* via conditioning on the observed variable *X*, which makes the two populations of agents comparable, as in a randomized study.

4 Three applications of the IV model

Two applications apply the IV model while the third one uses the Matching model.

4.1 Effect of risk management on firm value[4]

Reverse causality between firm hedging behavior and other firm financial variables is a crucial concern in risk management studies; it has been identified as the major source of inconsistency in the literature. To control for this endogeneity, Dionne and Mnasri (2018) studied the real effects of hedging using an instrumental variable approach applied to the essential heterogeneity model (Heckman et al., 2006). They controlled for biases related to selection on unobservables and self-selection in the estimation of the Marginal Treatment Effects (MTEs) of hedging choice on firm market value, risk and accounting performance.

[4] We do not study enterprise risk management (ERM) in this chapter. Hedging is a risk management strategy while ERM is a framework for risk management in an enterprise. On ERM, see the chapter by Hoyt and Liebengerg (2024) in this book.

They also estimated the Average Treatment Effects (ATEs), which can be interpreted as the mean of the MTEs. They estimated the hedging equation with the Probit model to reduce potential bias associated with unobserved heterogeneity.

In the last three decades, the risk management literature has been improved by data availability and improvements in theoretical research on corporate risk management. Mayers and Smith (1982) and Stulz (1984) were the first to build a hedging theory that incorporated the introduction of frictions into financial markets, and to show that market frictions (e.g., default costs, tax shields, agency costs, asymmetric information) enable firms to create value by hedging actively. The subsequent empirical literature extended the knowledge of hedging determinants (e.g., Tufano, 1996; Haushalter, 2000; Dionne and Garand, 2003; Adam and Fernando, 2006; Dionne and Triki, 2013). Other contributions in the literature focus on hedging value and risk implications for firms (e.g., Guay, 1999; Allayannis and Weston, 2001; Jin and Jorion, 2006). Empirical findings on the value implications of risk management are still fairly mixed and inconclusive. Methodological problems related to endogeneity of derivatives use and other firm decisions, sample selection, sample size, and the existence of other potential hedging mechanisms (e.g., operational hedging) are often blamed for this mixed empirical evidence.

To overcome the major source of inconsistency in the findings in the empirical literature (i.e., endogeneity), Dionne and Mnasri (2018) applied the instrumental variable approach with a model that considers essential heterogeneity inspired by the work of Heckman et al. (2006), which controls for the individual-specific unobserved heterogeneity in the estimation of marginal treatment effects of using high hedging ratios (i.e., upper quartile) versus low hedging ratios (i.e., lower quartile). Heckman et al. (2006) showed that the basic method of instrumental variables appears to be inappropriate when firms exhibit heterogeneous responses to treatment. In their application of the essential heterogeneity model, Dionne and Mnasri (2018) identified a credible instrument arising from the economic literature pertaining to the macroeconomic responses to crude oil price shocks, namely the Kilian (2009) index, which gives a measure of the demand for industrial commodities driven by the economic perspective.

Their evidence suggests that marginal firm financial value (marginal treatment effect, MTE), as measured by Tobin's *q*, is increasing with oil producers' propensity to hedge their oil production to a greater extent (i.e., upper quartile). This finding corroborates the stream in the literature that argues for the existence of a hedging premium for non-financial firms (Allayannis and Weston, 2001; Carter et al., 2006; Adam and Fernando, 2006; Pérez-Gonzalez and Yun, 2013). Consistent with the literature (e.g., Guay, 1999; Bartram et al., 2011), Dionne and Mnasri (2018) find that marginal firm riskiness, as measured by its systematic and idiosyncratic risks, is decreasing with oil producers' propensity to be high intensity rather than low intensity hedgers. Oil beta, representing firms' stock returns' sensitivity to fluctuations in oil prices, is decreasing with the propensity to hedge to larger extents, albeit with no statistical significance. Altogether, these findings suggest that any potential positive effects associated with oil hedging should translate into value enhancement for shareholders because of the decrease in the required cost of equity due to the lower riskiness of the oil producers, in particular lower systematic risk, as suggested by Gay et al. (2011). Dionne and Mnasri (2018) also found that the firm's marginal accounting performance, as measured by the return on equity, is lower for oil producers that are low intensive hedgers. Finally, the researchers obtained a significant average treatment effect (ATE) on firm value and risk for Tobin's *q* (positive), idiosyncratic risk (negative), and systematic risk (negative). ATE is not significant for accounting performance even if some firm types obtain significant MTE values.

Instrumental variable

For the choice of a candidate instrument, Dionne and Mnasri (2018) built on two previous studies that demonstrated a significant impact of oil market conditions (oil spot price and volatility) on oil hedging design in terms of maturity and type of derivatives (Mnasri et al., 2017, Dionne et al., 2018). They looked for an instrument that can explain the fluctuations of the real oil price and that cannot directly affect the value, riskiness and accounting performance of an oil producer. A large body of economic literature affirms that one of the most important fundamental factors that determines industrial commodity prices is demand pressures or shocks induced by real economic activity. Consequently, they chose the Kilian (2009) index as an instrument. This instrument measures the component of true global economic activity that drives demand for industrial commodities. This index is based on dry cargo (grain, crude oil, coal, iron ore, etc.) and single-voyage ocean freight rates, and captures demand shifts in global industrial commodity markets. The Kilian index, constructed monthly, accounts for fixed effects for different routes, commodities and ship sizes. It is also deflated with the US consumer price index and linearly detrended to remove the decrease in real terms over time of the dry cargo shipping cost. Kilian (2009) showed that aggregate shocks for industrial commodities cause long swings in real oil prices. This differs from the variations in the price of oil induced by oil market-specific supply shocks, which are more transitory. Oil market-specific supply shocks also differ from shocks related to shifts in the precautionary demand for oil, which arise from uncertainty about expected supply shortfalls relative to expected demand. Dionne and Mnasri (2018) calculated the changes in the Kilian (2009) index for each fiscal quarter in the sample. These changes in the index are calculated by taking the index's level at the end of the current fiscal quarter (i.e., at the end of the fiscal quarter's last month), minus its level at the end of the previous fiscal quarter. They showed that an increase in demand for industrial commodities is correlated with an increase in the prices of derivatives such as futures contracts. Consequently, oil hedging intensity should have a negative relationship with the Kilian index.

Essential heterogeneity model

The essential heterogeneity model usually begins with a Mincer–like equation (Mincer, 1974), as follows:

$$Y_{i,t} = \alpha + \beta D_{i,t} + \Delta X_{i,t} + u_{i,t}, \tag{15}$$

where $Y_{i,t}$ is the risk or value of oil producer *i* at the end of quarter *t,* and $D_{i,t}$ is the observed value of a dummy variable $D = (0,1)$ representing whether oil producer *i* uses low (0) or high (1) intensity hedging during quarter *t*. The control variables include a set of observable covariates $(X_{i,t})$. The term $u_{i,t}$ is an individual-specific error term and β represents the average return from using high intensity hedging (average treatment effect, ATE). Δ is a vector of parameters.

Two sources of bias could affect the estimates of β. The first is related to the standard problem of selection bias, when $D_{i,t}$ is correlated with $u_{i,t}$. This bias could be resolved by using an instrumental variable (IV) method, among others. This method may not be sufficient, however. The second source of bias occurs if the returns from using high intensity hedging vary across oil producers (i.e., β is random because of firm non-observed factors that can influence both the firm target and the hedging decision, such as governance of the firm or manager risk aversion), even after conditioning on observable characteristics leading to heterogeneous treatment effects. Moreover, oil producers make their hedging level choice (low versus high intensity) with at least partial knowledge of the expected idiosyncratic gains from this decision.

Heckman et al. (2006) developed an econometric methodology based on IVs to solve the problem of essential heterogeneity (i.e., β is correlated with $D_{i,t}$) in the estimation of MTEs. Their model is an extension of the LATE model developed by Imbens and Angrist (1994). MTE is a limit measure of LATE. Heckman et al.'s (2006) methodology is built on the Generalized Roy model, which is an example of treatment effect models for economic policy evaluation. The Generalized Roy model involves a joint estimation of an observed continuous outcome and its binary treatment.

We now present the linear and separable version of the essential heterogeneity model. Let (Y_0, Y_1) be the potential outcomes observed under the counterfactual states of treatment (Y_1) and no treatment (Y_0). These outcomes are supposed to depend linearly upon observed characteristics (X_{it}) and unobservable characteristics (U_0, U_1) as follows:

$$Y_1 = \alpha_1 + \beta + \Delta_1 X_{it} + U_1 \tag{16}$$

$$Y_0 = \alpha_0 + \Delta_0 X_{it} + U_0 \tag{17}$$

where β is the benefit related to the treatment $D_{i,t} = 1$.

The selection process is represented by $I_D = \gamma Z - V$, which depends on the observed values of the *Z* variables and an unobservable disturbance term (V) given *X*. The selection process, related to whether low or high intensity hedging is used, is linked to the observed outcome through the latent variable I_D, which gives the dummy variable *D* representing the treatment status:

$$D = \begin{cases} 1 \; if \; I_{D_{i,t}} > 0 \\ 0 \; if \; I_{D_{i,t}} \leq 0 \end{cases} \tag{18}$$

where the vector of observed *Z* variables includes instrumental variables Z_{IV} and all the components of $X_{i,t}$ in the outcome equation. The variable Z_{IV} satisfies the following constraints: $\text{Cov}(Z_{IV}, U_0) = 0$, $\text{Cov}(Z_{IV}, U_1) = 0$, and $\gamma_{IV} \neq 0$. The unobservable set of (U_0, U_1, V) is assumed to be statistically independent of *Z*, given *X*. In this application, one must first estimate the probability of participation in high intensity hedging or the propensity score and then analyze how this participation affects firm values and risks.

We can assume the joint normality of the outcome's unobservable components and decision equations $(U_0, U_1, V) \sim N(0, \Sigma)$, where Σ is the variance–covariance matrix of the three unobservable variables and $\sigma_{1V} = \text{Cov}(U_1, V)$, $\sigma_{0V} = \text{Cov}(U_0, V)$, and $\sigma_{VV} = 1$ following standard hypotheses. Under this parametric approach, the discrete choice model is a conventional probit with $V \sim N(0,1)$ where the propensity score $(P(z))$ is given by:

$$P(z) = \Pr(D = 1 \mid Z = z) = \Pr(\gamma z > V) = \Phi(\gamma z), \tag{19}$$

where $\Phi(\cdot)$ is the cumulative distribution of *V*, a standard normal variable. The term $P(z)$ denotes the selection probability of using high intensity hedging conditional on $Z = z$. We can therefore write:

$$\Phi(\gamma Z) > \Phi(V) \Leftrightarrow P(Z) > U_D \tag{20}$$

where

$$U_D = \Phi(V) \text{ and } P(Z) = \Phi(\gamma Z) = \Pr(D = 1 | \text{Z}).$$

The term U_D is a uniformly distributed random variable between zero and one representing different quantiles of the unobserved component V in the selection process. These two quantities, $P(Z)$ and U_D, play a crucial role in the essential heterogeneity model. The quantity $P(Z)$ could be interpreted as the probability of going into treatment and U_D interpreted as a measure of individual-specific resistance to undertaking treatment (or, alternatively, the propensity to not being treated as a high intensity hedger). In the application, the higher the $P(Z)$, the more the oil producer is induced to hedge its oil production to a larger extent due to Z. Conversely, the higher the U_D measures, the more resistant the oil producer is to using higher hedging extents due to a larger unobserved component. $P(Z)=U_D$ is thus the margin of indifference for oil producers that are indifferent between low and high intensity hedging.

The marginal treatment effects (MTEs) can be measured as follows:

$$MTE\left(X=x,U_X=u_X\right)=\left(\alpha_1+\beta-\alpha_0\right)+\left(\Delta_1-\Delta_0\right)x+\left(\lambda_{1V}-\lambda_{0V}\right)\Phi^{-1}\left(u_X\right) \quad (21)$$

Under the assumption of joint normality, λ_{1V} and λ_{0V} are the inverse Mills ratios coefficients used to reduce selection bias. They are estimated along with the other parameters in the two following equations:

$$E\left(\mathrm{Y|X}=x,D=1,P(Z)=p\right)=\alpha_1+\beta+\Delta_1x+\lambda_{1V}\left(-\frac{\phi\left(\Phi^{-1}(p)\right)}{p}\right) \quad (22)$$

$$E\left(\mathrm{Y|X}=x,D=0,P(Z)=p\right)=\alpha_0+\Delta_0x+\lambda_{0V}\left(\frac{\phi\left(\Phi^{-1}(p)\right)}{1-p}\right) \quad (23)$$

to obtain the $\widehat{MTE}$ values. Using the estimated propensity score:

$$\widehat{MTE}\left(X=x,U_X=u_X\right)=\widehat{\alpha_1+\beta}-\widehat{\alpha_0}+\left(\widehat{\Delta_1}-\widehat{\Delta_0}\right)x+\left(\hat{\lambda}_{1V}-\hat{\lambda}_{0V}\right)\Phi^{-1}\left(u_X\right). \quad (24)$$

Intuitively, how the MTE evolves over the range of U_D illustrates the heterogeneity in treatment effects among oil producers, that is, how the coefficient β is correlated with the treatment indicator D in (15). Equivalently, the estimated MTE shows how the increment in the marginal firm value, risk and accounting performance by going from choice 0 to choice 1 varies with different quantiles of the unobserved component V in the choice equation. Whether MTE increases or decreases with U_D tells us whether the coefficient β in (15) is negatively or positively correlated with the latent tendency of using high intensity hedging for oil production. Detailed estimation results, including firms' individual parameters (MTE), are presented by Dionne and Mnasri (2018). They clearly show that MTE varies between firms, meaning that the effect of hedging on firm value and risk is not homogenous between firms.

4.2 Asymmetric information in mergers and acquisitions

Dionne et al. (2015) studied the effect of asymmetric information in auctions. Their application concerns mergers and acquisitions of non-financial firms under asymmetric information. Their econometric analysis served to identify the causal effect on the premium of transaction value associated with the presence of blockholders. Does the presence of a blockholder on the board of a target firm affect the premium paid by the winner of the auction? This bidder should have more information on the target than the other potential bidders.

Econometric model

The authors postulate a linear relationship between the premium from a merger and acquisition and the candidate explanatory variables. They assume that more informed bidders (blockholders) pay a lower premium when they acquire a firm because other bidders know the blockholders have more information about the target. Consequently, they bid less actively because they do not want to pay an overly high price (overbidding). The presence of a blockholder on board is public information in the market since blockholders, in this study, hold more than 5% of capital stock value of the target. Let X_i denote a row vector that contains all available regressors, including the constant but excluding $Blockholders_i$. The model becomes:

$$Y_i = X_i\underline{\beta} + \beta_{15}Blockholders_i + \varepsilon_i \tag{25}$$

where $Y_i = Premium_i$ and $\underline{\beta} = (\beta_0, \dots \beta_{14})'$. The test for the null hypothesis (H_0) of no information asymmetry consists in verifying $\beta_{15} = 0$ against the alternative (H_1) $\beta_{15} < 0$. β_{15} measures the effect of blockholders' presence. In this application, β_{15} is assumed constant across firms in the absence of essential heterogeneity analysis.

It is straightforward to estimate this model by ordinary least squares. However, there are reasons to suspect that the blockholders dummy variable is correlated with unobservable factors in (25), in which case the ordinary least square estimate of β_{15} would be biased. For instance, blockholders may submit bids during the acquisition process because they want to be more active in a particular industry, and the current offer by a competitor may reduce this opportunity. This behavior is not observed by market participants. The endogeneity of $Blockholders_i$ would imply that $E(\varepsilon_i | X_i, Blockholders_i) \neq 0$. Below, we examine two approaches to deal with this endogeneity issue.

Treatment effect approach

The endogeneity of the $Blockholders_i$ variable can be addressed in line with the literature on the treatment effect (Heckman and Vytlacil, 2001). Indeed, the presence of blockholders in an auction may be viewed as a treatment that affects the distribution of the error term in (25) such that:

$$Y_i = X_i\underline{\beta} + \beta_{15} + \varepsilon_{i1} = Y_{i1},\ if\ Blockholders_i = 1, \tag{26}$$

and

$$Y_i = X_i\underline{\beta} + \varepsilon_{i2} = Y_{i2},\ if\ Blockholders_i = 0, \tag{27}$$

where ε_{i1} and ε_{i2} are errors with potentially distinct distributions. One may further ask: Is there any hidden self-selection effect in the process that drives the presence of blockholders in auctions? Indeed, blockholders may be keener to attend auctions where their informational advantage is higher. As a result, their ability to lower the premium would be higher for the auctions in which they participate to win than in other auctions.

According to the treatment effect formulation, the error term of (25) is given by:

$$\varepsilon_i = \varepsilon_{i1} Blockholders_i + (1 - Blockholders_i)\varepsilon_{i2}. \tag{28}$$

To estimate the coefficients of (25) by OLS, the initial set of predictors is augmented with a set of instrumental variables denoted Z_i. Consequently, the premium can be represented as:

$$Y_i = X_i\underline{\beta} + \beta_{15}Blockholders_i + E(\varepsilon_i|X_i, Z_i, Blockholders_i) + \tilde{\varepsilon}_i \tag{29}$$

where the new error term satisfies $\tilde{\varepsilon}_i = \varepsilon_i - E(\varepsilon_i|X_i, Z_i, Blockholders_i)$ so that $E(\tilde{\varepsilon}_i|X_i, Z_i, Blockholders_i) = 0$. When estimating (29), we can assume the existence of a latent variable k_i such that $Blockholders_i = 0$ if $k_i \leq 0$ and $Blockholders_i = 1$ if $k_i > 0$. This latent variable further satisfies:

$$k_i = X_i\underline{\delta} + Z_i\underline{\gamma} + e_i \tag{30}$$

where $e_i \sim N(0,1)$ and $\underline{\delta}$ and $\underline{\gamma}$ are vectors of parameters.

Equation (30) was estimated by a Probit model using three variables as instruments: (i) Intrastate, (ii) Regulated industry and (iii) an interaction variable between Intrastate and Performance of the target. The first variable comes from the study by Kang and Kim (2008). They found that blockholders prefer targets in the same state because proximity reduces the transaction costs, yielding higher returns. They also showed that the monitoring of intrastate firms is more valuable for targets that have poor past performance. Blockholders are consequently more likely to buy shares in intrastate underperforming firms to better exploit their informational advantage. Blockholders may be more present in poorly performing targets of their state because they have a higher probability of obtaining long-run value by exploiting asymmetric information with other competitors. The other variable controls for the fact that the industry of the target is regulated. Blockholders should better exploit their informational advantage for these firms because they are more knowledgeable about the different laws regulating the target. Considering only auctions where blockholders are present, the effect of their presence compared to the outcome if they were absent is given by:

$$E(Y_{i1} - Y_{i2}|X_i, Z_i, Blockholders_i = 1) = \beta_{15} + \lambda_1 \frac{\varphi(\hat{\eta}_i)}{\Phi(\hat{\eta}_i)}, \tag{31}$$

where λ_1 is the coefficient of the inverse Mill ratio (IMR). Similarly, considering only auctions where blockholders are not present, their absence would have induced the following effect:

$$E(Y_{i1} - Y_{i2}|X_i, Z_i, Blockholders_i = 0) = \beta_{15} - \lambda_2 \frac{\varphi(\hat{\eta}_i)}{1-\Phi(\hat{\eta}_i)}. \quad (32)$$

Accordingly, the average treatment effect (ATE) of the presence of blockholders on the premium is:

$$E(Y_{i1} - Y_{i2}|X_i, Z_i) = \beta_{15}.$$

Finally, the equation that lets us estimate all parameters without bias is given by:

$$Premium_i = X_i\underline{\beta} + \beta_{15}Blockholders_i + IMR_i\underline{\lambda} + \tilde{\varepsilon}_i \quad (33)$$

where $\underline{\lambda} = (\lambda_1, \lambda_2)$, $IMR_i = [IMR_{1i}, IMR_{0i}]$, $IMR_{1i} = \frac{\varphi(\hat{\eta}_i)}{\Phi(\hat{\eta}_i)}Blockholders_i$, and $IMR_{0i} = \frac{-\varphi(\hat{\eta}_i)}{1-\Phi(\hat{\eta}_i)}(1 - Blockholders_i)$.

Two estimation approaches

Given that $Blockholders_i$ is a binary variable, two approaches can be considered to proxy its expectation conditional on X_i and Z_i. The first approach relies on the linear probability model (LPM):

$$Blockholders_i = X_i\underline{\delta} + Z_i\underline{\gamma} + e_i, \quad (34)$$

from which the fitted values are deduced as $\widehat{Blockholders}_i = X_i\underline{\hat{\delta}} + Z_i\underline{\hat{\gamma}}$.

The second approach is based on the Probit model presented in the previous subsection, from which fitted values are deduced as $\widehat{Blockholders}_i = \Phi(\hat{\eta}_i)$. Either the LPM or the Probit model can be a good proxy for $E(Blockholders_i = 1|X_i, Z_i)$ (see, however, Heckman and Pinto, 2022). Hence, the first-stage functional form does not affect the consistency of the second-stage estimates. The second stage estimating equation is:

$$Y_i = X_i\underline{\beta} + \beta_{15}\widehat{Blockholders}_i + \hat{\varepsilon}_i. \quad (35)$$

Two formal tests are done to assess the validity of the results: Sargan's over-identifying restrictions test for the exogeneity of the instruments (H0: the instruments are exogenous) and the Durbin-Wu-Hausman test for the relevance of the instrumental variable method (H0: the *Blockholders* variable is exogenous). These tests are performed only within the 2SLS approach that relies on the LPM (linear) at the first stage of the estimation. The tests cannot be used with the Probit (non-linear) model.

Dionne et al. (2001) (see also Blundell and Smith, 1989) propose the following extension of Equation (35) to make it robust to nonlinearity and misspecification within the framework that employs a Probit at first stage:

$$Y_i = X_i\underline{\beta} + \tilde{\beta}_{15}\Phi(\hat{\eta}_i) + \beta_{15}Blockholders_i + \mu_i. \tag{36}$$

By adding and subtracting $Blockholders_i - \Phi(\hat{\eta}_i)$ to $\Phi(\hat{\eta}_i)$, Equation (36) becomes:

$$Y_i = X_i\underline{\beta} + \left(\beta_{15} + \tilde{\beta}_{15}\right)Blockholders_i - \tilde{\beta}_{15}\left(Blockholders_i - \Phi(\hat{\eta}_i)\right) + \mu_i. \tag{37}$$

$Blockholders_i$ is endogenous if $\underline{\delta}$ in (34) is significantly different from zero. By comparing (35) and (36), we see that $Blockholders_i$ is endogenous if $\tilde{\beta}_{15}$ is significantly different from zero in the model of Dionne et al. (2001). In that event, the (unbiased) coefficient of $Blockholders_i$ is equal to $\beta_{15} + \tilde{\beta}_{15}$ as indicated by (37). The estimation results clearly show that more informed bidders pay a lower premium in mergers and acquisitions (Dionne et al., 2015). The three instrumental variables were valid instruments, meaning that the presence of blockholders is a causal variable to explain the lower premium paid by the blockholders in the presence of asymmetric information.

The estimation results, which consist of the results for the standard OLS, the treatment effect approach, the 2SLS approaches (2SLS-LPM and 2SLS-Probit), and Dionne et al. (2001) are presented in Table 5 of the study.

The results from all models clearly support the hypothesis that information asymmetry between potential buyers significantly influences the premium paid during an acquisition. This result is consistent with the theoretical and empirical study (Dionne et al., 2009) showing that information asymmetry between the participants and even adverse selection influences the equilibrium price of an auction.

4.3 CDS central clearing

Akari et al. (2021) revisited the impact of the voluntary central clearing scheme on the CDS market. Central clearing is intended to improve transparency in the derivatives market by reducing asymmetric information between market participants in Over the Counter Transactions (OTC). This article is part of the ongoing research on the impact of introducing a central counterparty (CCP) that stands between buyers and sellers of default protection in the CDS market.

Akari et al. (2021) applied the dynamic propensity-score matching methodology combined with difference-in-differences (DID). Their empirical findings show that central clearing results in a small, but significant, increase (estimated at 19 bps) in CDS spreads, while there is no evidence of an associated improvement in CDS market liquidity and trading activity or of a deterioration in the default risk of the underlying bond. These results suggest that the increase in CDS spreads can be mainly attributed to a reduction in CDS counterparty risk in the central market because CDS buyers are protected against issuer default. The spread, paid by risk adverse investors, is used to finance part of the CCP activities.

Their approach aims at eliminating the selection bias in causal analysis. First, they used a dynamic matching technique, relying on dynamic firm data just prior to the move to central clearing, instead of using a fixed estimation period to match all the firms. Second, they estimated a DID model including time and firm fixed effects.

In order to study the impact of central clearing, they compared the spreads of single-name CDS contracts in two groups of firms, namely cleared reference entities that are members of the clearinghouse and non-cleared reference entities that are in the OTC market; this comparison is undertaken before and after adhesion to the CCP, in a DID framework.

The classical two-by-two design uses data from a treatment group and from a control group, measured at two different dates: before treatment and after treatment. This methodology is flexible and can be generalized to the case of multiple groups and multiple time periods (Bertrand et al., 2004; Imbens and Wooldridge 2009; Gormley and Matsa 2011). In this application, since the authors were dealing with multiple treatment (clearing) dates, they opted for a DID framework with firm and time fixed effects.

To apply DID with Matching, one needs to form a treatment and a control group containing firms that have similar characteristics just before the treatment event. The first step consists of constructing a sample of candidate treatment and control entities, and computing their propensity scores on the basis of pre-clearing characteristics. Specifically, the authors considered the 29 clearing dates as the various possible times for adhering to a CCP. These treatment dates can be interpreted as hypothetical events for the control group. Each non-cleared firm thus generates up to 29 firm-event entities. The sample also contains the cleared firms, paired with the event corresponding to their clearing date.

The Probit model was then estimated by using the sample of cleared and non-cleared firm-event entities and the corresponding observable variables that are relevant to clearinghouses:

$$\Pr\left(Y=1\middle|X\right)=\Phi\left(X\beta\right), \tag{38}$$

where Y is a binary random variable that equals 1 if the firm is centrally cleared and 0 otherwise. Φ is the standard normal cumulative distribution function, X is the vector of regressors that influence the outcome Y, and β is a vector of parameters. The vector β is obtained by maximum likelihood and is used to estimate the probability, for each firm-event entity, of being accepted for central clearing. This probability is the propensity score associated to a combination of firm characteristics and a possible clearing date.

The second step consists of matching cleared and non-cleared entities on the basis of the propensity scores. One can match with replacement each cleared firm with its closest neighbor from the group of non-cleared firm-event entities. The final sample is then composed of matched firm-event entities.

A DID regression to the matched sample can be applied to test for the presence of statistically significant impact factors such as liquidity. More specifically, to isolate the effect of central clearing on a given factor, one can estimate the following DID equation:

$$Factor_{i,t} = \beta_0 + \beta_1 Cleared_{i,t} + \alpha_i + \gamma_t + \varepsilon_{i,t}, \tag{39}$$

where subscript i denotes a firm-event entity and subscript t denotes a date in the event window. The dependent variable $Factor_{i,t}$ takes various definitions in order to investigate the impact of central clearing on CDS spreads, liquidity, and trading activity, as well as on bond default spreads. The main explanatory variable $Cleared_{i,t}$ is a binary variable that indicates whether the reference entity i is centrally cleared or not on date t. This variable is the equivalent of the interaction term in the classic two-by-two DID design because Matching eliminated the differences between units before the events. The treatment effect is given by the corresponding coefficient β_1. The fixed effects of the generalized DID setting help control for unobserved heterogeneity across time and reference entities, thereby alleviating concerns about any omitted variables that might affect both groups in the same way. The firm fixed effect, α_i, captures differences across firms that are constant over time, while the time fixed effect, γ_t, captures differences over time that are common to all firms. The authors deliberately did not control for specific time-varying variables to avoid confounding estimates of β_1, since these variables might also be affected by the move to central clearing.

5 Measurement of residual asymmetric information in insurance data

The objective of this section is to present various tests for the presence of residual asymmetric information in insurance markets. From the theoretical literature (Dionne et al., 2024; Picard, 2024; Parra and Winter, 2024; Crocker et al., 2024), we know that the potential presence of asymmetric information between insured and insurers regarding individual risks motivates partial insurance, risk classification, long-term contracts, and auditing of claims. It is also well known from the insurance literature that risk classification is due, in part, to asymmetric information between the insurer and the insured (Hoy, 1982; Crocker and Snow, 1985, 1986; Dionne and Rothschild, 2014). Full efficiency in risk classification should separate individual risks and generate different actuarial insurance premiums that reflect these risks. This means there should not be any residual asymmetric information between the insurer and the insured inside the risk classes when risk classification is costless and not regulated. In the presence of proportional transaction costs, partial insurance should be the optimal contract, but there should be no correlation between insurance coverage and individual risk. However, in the real world of insurance contracting, there may be numerous constraints that limit efficiency in risk classification. Incentive contracting to obtain more information on individual risk in different risk classes thus becomes important, and the empirical question is how efficiently this mechanism reduces asymmetric information in insurers' portfolios.

Cohen and Siegelman (2010) present a survey of empirical studies of adverse selection in insurance markets. They argued that the coverage-risk correlation is particular to each market. Accordingly, the presence of a significant coverage-risk correlation has different meanings in different markets, and even in different risk pools in a given market, depending on the type of insured service, the participants' characteristics, institutional factors, and regulation. This means that when testing for the presence of residual asymmetric information, one must also control for these factors. What characteristics and factors explain the absence of coverage-risk correlation in automobile insurance markets? Some studies using the conditional correlation approach on cross-sectional data found evidence of asymmetric information (Dahlby, 1983, 1992; Devlin, 2002; Puelz and Snow, 1994; Richaudeau, 1999; Cohen, 2005) while others did not (Chiappori and Salanié, 2000; Dionne et al., 2001; Saito, 2006).[5] One major criticism of the conditional correlation approach with cross-sectional data is that it does not allow one to conclude on causality.

The first empirical question in insurance markets can be summarized as follows: Is there any residual causal effect between chosen insurance coverage and risk within risk classes? The second question is how to identify which information problem remains when the first test rejects the null hypothesis that there is no residual information problem. This step is important for the insurer because it must implement the appropriate instruments to improve resource allocation. A deductible efficiently reduces *ex ante* moral hazard, but not necessarily *ex post* moral hazard because often, the accident has already occurred when the

[5] Other recent studies include those by Kim et al., 2009; Olivella and Vera-Hernández, 2013; Dardanoni et al., 2018; Geyer et al., 2020.

action is taken. A high deductible may even have an adverse effect and encourage accident cost building. As is well known in the empirical literature, a positive correlation between insurance coverage and risk is a necessary condition for the presence of asymmetric residual information, but it does not shed light on the nature of the information problem. The third question is how improving the contracts can reduce the negative impact of asymmetric information on resource allocation. These resource allocation objectives must take into account other issues such as risk aversion, fairness, and accessibility of insurance. This last issue is particularly important in many insurance markets. A decrease in insurance coverage may reduce *ex ante* moral hazard because it exposes the insured person to risk, but it also significantly reduces accessibility to insurance protection for risky and poor people who are not always responsible for their risk type and financial conditions.

Econometricians analyze two types of information when studying insurers' data (Boyer and Dionne, 1989; Dionne and Vanasse, 1989, 1992; Chiappori, 1994; Puelz and Snow, 1994; Gouriéroux, 1999; Richaudeau, 1999; Dionne and Ghali, 2005; Saito, 2006; Dionne et al., 2006; Lee, 2013; Spindler, 2014; Dionne and Harrington, 2014; Zavadil, 2015; Rowell et al., 2017a, 2017b). The first type contains variables that are observable by both parties to the insurance contract. Risk classification variables are one example. Econometricians/insurers combine these variables to create risk classes when estimating accident distributions. Observed variables can be used to make estimates conditional on the risk classes or within the risk classes. The second type of information is related to what is not observed by the insurer or the econometrician during contract duration and at contract renegotiations, but can explain the insured's choice of contracts or safety in driving. If we limit our interpretation to asymmetric information (either moral hazard, adverse selection or both), we can test the conditional residual presence of asymmetric information in an insurer's portfolio or look for a correlation between the contract coverage and the realization of the risk variable during a contract period. Two parametric tests have been proposed in the literature (Chiappori and Salanié, 2000; Dionne et al., 2001; see Chiappori and Salanié, 2013, for a detailed analysis). One parametric test (Dionne at al. 2001) estimates the following relationship:

$$y_i = g\big(\alpha + \beta X_i + \gamma d_i + \delta E(d_i | X_i)\big) + \varepsilon_i, \tag{40}$$

where y_i is the contract choice by individual i (level of deductible, for example), X_i is a vector of control variables such as the observable characteristics used in risk classification and control variables for risk aversion, β is a vector of parameters to be estimated d_i is the realization of the random variable observed at the end of the contract period (accident or not, for example), $E(d_i|X_i)$ is the conditional expected value of the random variable obtained from the estimation of the accident distribution, and ε_i is the residual of the regression. A positive sign is usually anticipated for the coefficient d_i when residual asymmetric information remains (higher coverage is related to more accidents or higher risk). The seminal theories of Rothschild and Stiglitz (1976) and Wilson (1977) strongly predict that such a correlation should be observed in the data in the presence of adverse selection, while Holmström (1979) and Shavell (1979) strongly predict that the correlation is due to moral hazard. Note that the dependent variable in the above regression can be the risk variable d_i while coverage y_i is an independent variable. The presence of the variable d_i is not necessarily exogenous in equation (40). It is better to instrument this variable in order to obtain a causality result, as shown in Dionne et al. (2009) and Rowell et al. (2017a, 2017b).

The presence of $E(d_i|X_i)$ is necessary to control for specification errors (missing variables) or for potential non-linearity not modeled in the equation. Without this control, the coefficient (γ) of d_i can be significant for reasons other than the presence of residual asymmetric information in the risk classes.

If the coefficient of d_i is not significant, one can reject the presence of residual asymmetric information in the risk classes when all other factors are well controlled. This does not mean that there was no asymmetric information in this market; rather, it means that the insurers risk classification system eliminated asymmetric information efficiently, and that there is no *residual* asymmetric information within the risk classes. In other words, when risk classification is done properly, it is not necessary to choose the insurance contract form within the risk classes to reduce asymmetric information. So the observed presence of a deductible must be explained by other factors in the insurance market such as transaction costs.

An equivalent parametric model was proposed by Chiappori and Salanié (2000). Here, two equations are estimated simultaneously, one for contract choice and the other for accident distribution. An example is the bivariate probit model:

$$y_i = f(x_i, \beta) + \varepsilon_i \tag{41}$$

$$d_i = g(x_i, \beta) + \eta_i. \tag{42}$$

The test consists in verifying whether there is dependence between the residuals of the two equations. An absence of conditional correlation is interpreted as an absence of residual asymmetric information in the data. The authors performed an additional non-parametric test that is independent of the functional forms of the above model. It is based on a Chi-square test of independence. However this test seems to be limited to discrete variables, contrary to the two parametric tests presented above (see Su and Spindler, 2013, for a longer discussion).

Many extensions of these models have been presented in the literature. Chiappori and Salanié (2013) specified conditions to obtain robustness of the test when insured may have different degrees of risk aversion. They showed that if insurers maximize profits in competitive markets, the results of the above test are robust to heterogeneity in preferences. Such robustness is less evident in non-competitive insurance markets.[6]

Fang et al. (2008) did not reject asymmetric information in the medical insurance market, nor did they find evidence of adverse selection. Their results are consistent with multidimensional private information along with advantageous selection (De Meza and Webb, 2001). They obtained a *negative* correlation between risk and insurance coverage. Risk aversion is not a source of advantageous selection in their data. The significant sources are income, education, longevity expectations, financial planning horizons, and, most importantly, cognitive ability, (see also Finkelstein and McGarry, 2006, on this issue).

To separate moral hazard from adverse selection, econometricians need a supplementary step. An additional market relationship can be estimated to look for adverse selection

[6] Maliar and Salanié (2022) showed that adding deep learning to the test model does not affect the correlation results obtained by Chiappori and Salanié (2000).

(conditional on the fact that the null hypothesis of no asymmetric information was rejected), as Dionne et al. (2009) did for auctions where moral hazard is not a significant problem.

In insurance markets, dynamic data are often available. Time adds an additional degree of freedom to test for asymmetric information (Dionne and Lasserre, 1985, 1987; Dionne and Vanasse, 1989, 1992; D'Arcy and Doherty, 1990; Dionne and Doherty, 1994; Chiappori et al., 1994; Hendel and Lizzeri, 2003). This information can be used in many insurance markets where past experience information is available and when usage is possible. For ethical reasons, this information is not often utilized on an individual basis in health insurance and for bodily injury insurance in many countries. Experience rating works at two levels in insurance. Past accidents implicitly reflect unobservable characteristics of the insured (adverse selection) and introduce additional incentives for prevention (moral hazard). Experience rating can therefore directly mitigate problems of adverse selection and moral hazard, which often hinder risk allocation in the insurance market.

Experience rating not only provides additional information on risk but may also play an important role in the dynamic relationship between policyholders' insurance claims and contract choice. The theoretical literature on repeated insurance contracting over time clearly indicates that these features may help overcome problems of moral hazard when risks known to the policyholder (endogenous) are unobservable by the insurer (moral hazard, Parra and Winter, 2024) or when exogenous characteristics are unobservable (adverse selection, Dionne et al., 2024). Contract choice is influenced by the evolution of the premium, which is closely linked to the insured's risk or past experience. Because increased insurance coverage tends to lower the expected cost of accidents for the insured, incentives for safe behavior are weakened for all risks. Under experience rating, the subsequent rise in accidents increases the marginal costs of future accidents when experience rating is taken into account. Experience rating may therefore offset the disincentive effect created by single-period insurance coverage.

Abbring et al. (2003) applied a multi-period incentive mechanism by focusing on the dynamics of claims, but not on the dynamics of contract choice (because of data limitations). Proposing specific assumptions about the wealth effects of accidents to policyholders who differ only in their claim records (thus their experience rating), their model predicts that subjects with the worst claims records should try harder to increase safety, and thereby, ceteris paribus, file fewer claims in the future. However, their data do not support the presence of moral hazard. Dionne et al. (2011) extended their model and did not reject the presence of moral hazard, using a different data set. The potential presence of adverse selection in their data was not a real problem because all drivers must be insured for bodily injuries.

Dionne et al. (2013a) showed that failure to detect residual asymmetric information, and more specifically, moral hazard and adverse selection in insurance data, is due to the failure of previous econometric approaches to model the dynamic relationship between contract choice and claims adequately and simultaneously when looking at experience rating. Intuitively, because there are at least two potential information problems in the data, an additional relationship to the correlation between risk and insurance coverage is necessary to test for the causality between risk and insurance coverage. Using a unique longitudinal survey of policyholders from France, they propose a methodology to disentangle the historical pathways in claims and premiums. They explained how causality tests can be used to differentiate moral hazard from asymmetric learning (and eventually adverse selection). They did not reject moral hazard for a given group of policyholders, and did not reject asymmetric learning for younger drivers. The empirical methodologies of Dionne et al. (2011) and Dionne et al. (2013a) are reviewed in detail below, as well the recent study of Dionne and Liu (2021) that applied a difference-in-differences (DID) approach to separate moral hazard from adverse selection and evaluated the introduction of a new regulation on insurance pricing in China.

6 Causal tests for moral hazard in the automobile insurance market

6.1 Moral hazard as a function of accumulated demerit points

We now analyze moral hazard as a function of demerit points. Because no-fault environments are common in North America, traffic violations are events likely to be used in insurance experience rating schemes. In Quebec, the public insurer in charge of the compensation of bodily injuries uses an experience rating scheme based on demerit points.[7] The same public enterprise is also in charge of the point-record license system. Dionne et al. (2011) show that the new insurance pricing scheme based on demerit points introduced in 1992 reduced the number of traffic violations by 15%. They also determined that there is residual *ex ante* moral hazard in road safety management. The discussion below focuses on the methodology they developed for obtaining this result.

The methodology extends the empirical model of Abbring et al. (2003). Over time, a driver's observed demerit points informs on two effects: an unobserved heterogeneity effect and an incentive effect. Drivers with more demerit points accumulated during a period are riskier with respect to hidden features in risk distributions. Hence, unobserved heterogeneity is a form of risk reassessment in the sense that those who accumulate demerit points represent higher risks over time. This time effect is in the opposite direction of the incentive effect. For the incentive effects, accumulating demerit points should increase the incentive for safe driving to reduce the probability of receiving a higher penalty, such as losing the driving license for a certain period.

The model proposed by Dionne et al. (2011) tests for an increasing link between traffic violations and the number of accumulated demerit points over time. Rejecting the positive link is necessary to verify the presence of moral hazard. They estimated the following hazard function (Cox, 1972):

$$\lambda_{it} = \exp\left(x_{it}\beta\right) + g\left(adp_{it}\right) \times h\left(c_{it}\right) \tag{43}$$

[7] On point-record driver's licenses, see the study by Bourgeon and Picard (2007).

where λ_{it} is the hazard function for driver *i* at date *t*, x_{it} is a vector of control variables, β represents the corresponding coefficients, adp_{it} is the number of demerit points accumulated over the two previous years at time *t*, and c_{it} is contract time at date *t*.

In the absence of moral hazard, *g* should be increasing because of unobserved heterogeneity. They found that *g* is decreasing when drivers have accumulated more than seven demerit points. This means that beyond seven demerit points, drivers become safer because they do not want to lose their driver's license when they have accumulated 15 demerit points. This is evidence of the presence of moral hazard in the data: these drivers were negligent when the accumulated record was below seven demerit points (see Pinquet, 2024, and Dionne et al., 2013b, for more details).

6.2 Separating moral hazard from learning and adverse selection with dynamic data

To separate learning leading to adverse selection (asymmetric learning) from moral hazard, Dionne et al. (2013a) considered the case where information on contracts and accidents is available over time in the form of panel data. They exploited dynamics in accidents and insurance coverage while controlling for dynamic selection due to unobserved heterogeneity. They constructed two additional tests based on changes in insurance coverage. Coupled with the negative occurrence test of Abbring et al. (2003) and Dionne et al. (2011), their two additional tests allowed them to separate moral hazard from asymmetric learning (which should become adverse selection in the long run).

They analyzed the identification of asymmetric learning and moral hazard within the context of a tractable structural dynamic insurance model. From the solution of their theoretical model, they simulated a panel of drivers behaving under different information regimes or data generating processes (with or without both moral hazard and asymmetric learning). They validated their empirical tests on simulated data generated from these different information regimes. They then applied these tests to longitudinal data on accidents, contract choice and experience rating for the 1995-1997 period in France (Dionne, 2001). They found no evidence of information problems among experienced drivers (more than 15 years of experience). For drivers with less than 15 years of experience, they found strong evidence of moral hazard but little evidence of asymmetric learning. They obtain evidence of asymmetric learning, despite the small sample size, when focusing on drivers with less than five years of experience. To obtain these results, they estimated the following model.

They considered a joint parametric model for the probabilities of accidents and contract choice. Each equation corresponds to a dynamic binary choice model with pre-determined regressors and an error component structure. The error component structure is important given the likelihood of serial correlation in contract and accident outcomes. They used the solution proposed by Wooldridge (2010) to take the potential left censoring effect into account. More specifically, the process for accidents is specified as:

$$\begin{gathered} n_{it} = I(x_{it}\beta_n + \phi_{nd} d_{it-1} + \phi_{nn} n_{it-1} + \phi_{nb} b_{it} + \varepsilon_{n,it} > 0) \\ i = 1,...,N, t = 1,...,T \end{gathered} \tag{44}$$

where $\varepsilon_{n,it}$ has an error component structure $\varepsilon_{n,it} = \alpha_{ni} + v_{n,it}$, n_{it} is a binary variable for the number of accidents of individual *i* at time *t*, d_{it-1} is the individual's contract choice in period $t-1$, n_{it-1} is the number of accidents in period $t-1$, and b_{it} is the bonus-malus score at period *t*. The presence of moral hazard would be confirmed by a positive sign for ϕ_{nd} (more insurance coverage-more accidents) and a negative sign for ϕ_{nb} (a higher malus creates more incentives for safe driving, similar to the test presented in the previous section with accumulated demerit points.) Here a high malus means an accumulation of accidents over the previous periods. Dionne et al. (2013a) specify a similar equation for contract choice:

$$\begin{gathered} d_{it} = I(x_{it}\beta_d + \phi_{dd} d_{it-1} + \phi_{dn} n_{it-1} + \phi_{db} b_{it} + \varepsilon_{d,it} > 0) \\ i = 1,...,N, t = 1,...,T \end{gathered} \tag{45}$$

where again $\varepsilon_{d,it} = \alpha_{di} + v_{d,it}$. The asymmetric learning test is a test of whether an accident (not necessarily a claim) in the last period, conditional on the bonus-malus, leads to an increase in coverage this period. In this model, the insured observe all accidents while the insurer observes only the claims. Drivers thus learn that they are riskier than anticipated and increase their insurance coverage accordingly. It is a test of whether $\phi_{dn} > 0$ or not.

6.3 Separating moral hazard from adverse selection with a DID model

We now consider the DID approach to separate moral hazard from adverse selection by reviewing the study of Dionne and Liu (2021) (see also Abay, 2018). In this study, the authors analyzed the effect of a public reform of the Chinese bonus-malus system on road safety and showed how the observed variation in accident rates could be attributed to a reduction in moral hazard. They had access to panel data from two cities, one affected by the reform (treatment group) and one not affected (control group).

DID matching estimators

One potential concern is that the control vehicles may differ from the treated vehicles along most of the observable dimensions. The authors use the propensity score matching methodology (PSM) to balance the two groups in terms of observables. The goal of the reform was to improve road safety (fewer accidents) by introducing more incentives for safe driving in the pricing of auto insurance. They first estimated the propensity score or the likelihood of being treated by using a standard logit model. Their results affirmed that they obtained covariate balances between the matched treated and control groups for most of the dimensions they used. Potential adverse selection was thus eliminated. They then estimated DID matching estimators and determined that the number of accidents in the treated city dropped after the two-stage reform.

DID with parallel trends

The DID approach allows analysts to test the dynamics of the treatment effect induced by each of the two stages of the reform on road safety. A relative decline in claim frequency was expected in the treated group compared with the control group after both stages of the reform. Equation (46) presents the basic regression equation:

$$Claims_{it} = \alpha + \sum_{s=1}^{2} \beta_s treated_i \times \text{post-event}_{st} + \kappa_i + \eta_t + \varepsilon_{it} \tag{46}$$

where $s=1$ and $s=2$ denote the first-stage and second-stage reform respectively. β_1 and β_2 are the main parameters of interest: they evaluate the differential effects of the two-stage reform at each stage across the treated and control groups. $Claims_{it}$ measures claim frequency for vehicle *i* in year *t*. α is a constant term. $treated_i=1$ when the policy is issued in the treated city, and 0 otherwise. $treated_i=1$ and $\text{post-event}_{2t}=1$ when the due date of the policy is after the first-stage and second-stage reform respectively, and 0 otherwise. κ_i is the vehicle fixed effects, the proxy for driver fixed effects, which controls for vehicle/driver-level time-invariant heterogeneity. η_t is the year fixed effects, which accounts for the common aggregate shocks.

In addition, to verify the parallel trends in claim frequency across these two groups before the reform (no difference in claims variations), Dionne and Liu (2021) estimated the time-varying effects of the reform by year using the following distributed lag model in Equation (47).

$$Claims_{it}=\alpha+\sum_{y=-1}^{2}\beta_y treated_i\times year_{yt}+\kappa_i+\eta_t+\varepsilon_{it} \tag{47}$$

where $year_{yt}$ represents a set of 4 yearly dummies from one year before the reform until two years after the first-stage reform year. β_y measures the differential trend in claim frequency across the treated and control groups during the pre-reform period. The analysis was repeated with semester and quarterly observations. In all cases, the authors did not reject parallel trends between claims in the two cities before the reform, a necessary step to consolidate the DID methodology. Their estimation of Equation (46) confirmed the presence of moral hazard in the data. Adverse selection was ruled out by showing that there was no significant differences between drivers' and cars' characteristics between the two cities before and after the reform.

Figure 1 shows the evolution of accident rates during the period of analysis. We observe a clear difference between the two frequencies after the two reforms. The new pricing scheme of automobile insurance reduced the accident rates in the treated city, which is attributed to a reduction in ex-ante moral hazard. Figure 2 does not reject the presence of parallel trends between accident distributions before the first reform in the two pre-reform periods (–2, –1) and confirms the differences in accident rates after the reform (starting in period 3).

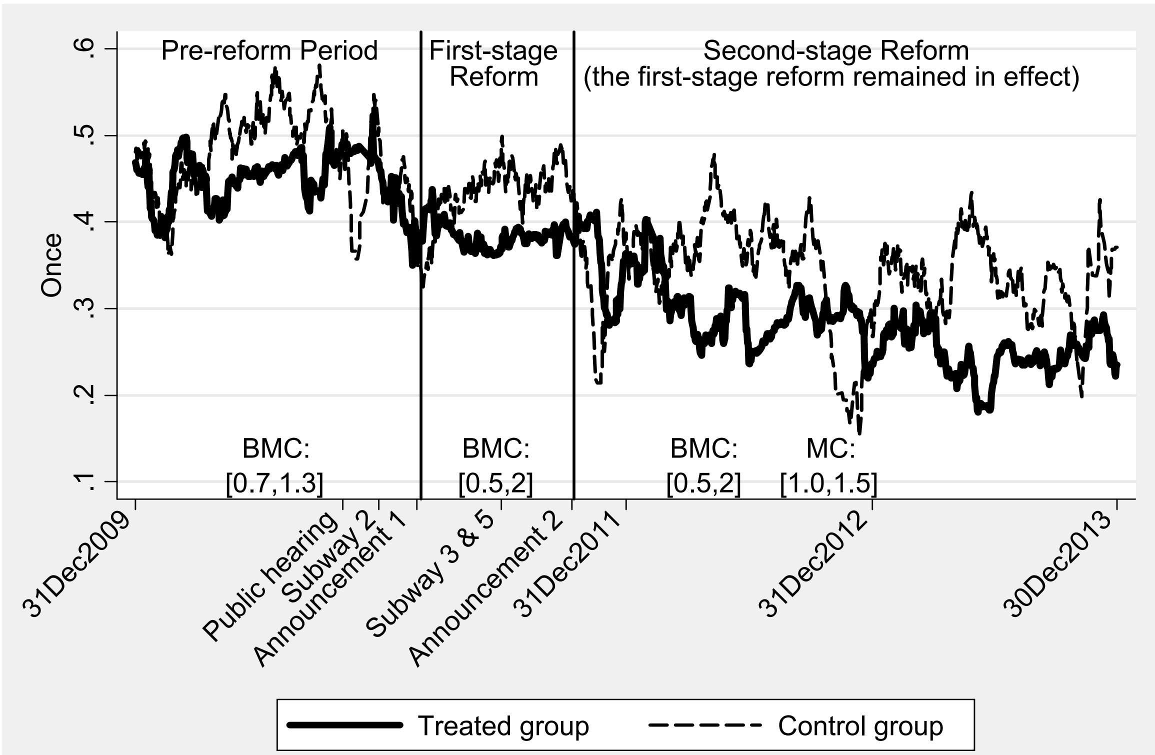


Figure 1 – Evolution of two-stage reform and claim frequency by group and period

Source: Dionne and Liu (2021)

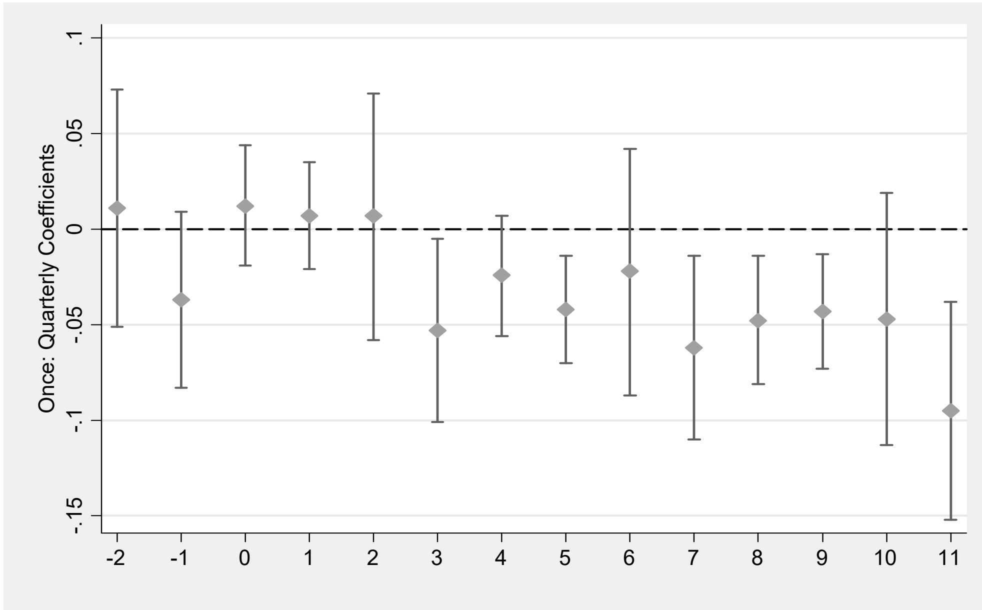


Figure 2 – Quarterly estimates of the effects of insurance incentives on road safety with two pre-reform periods

Source: Dionne and Liu (2021)

7 Causality analysis based on the generalized method of moments

7.1 Econometric model

Desjardins et al. (2022) studied the causal reciprocal relationship between liquidity creation and reinsurance demand. One objective of their research was to examine the reciprocal relationship between reinsurers' liquidity creation and their demand for reinsurance. One of the most efficient models that can be used to evaluate this reciprocal relation is the structural equations model (Low and Meghir, 2017). To evaluate the reciprocal relation between reinsurance demand and liquidity creation with a structural equations model, Desjardins et al. (2022) specified a dynamic data panel incorporating unobserved heterogeneity. In this model, the lagged value of liquidity creation was added in the equation of reinsurance demand as one of the key explanatory variables, and the lagged reinsurance demand variable was added in the equation of liquidity creation as an explanatory variable. This specification of the model, where the parameters associated with lagged variables reflect a causal link that takes some time to become effective. In fact, most of insurers' strategic decisions, such as investments (liquidity creation) and reinsurance demand (risk management), are generally made by the board of directors once a year and may take several months to materialize. The decisions are very unlikely to produce causal effects during the same year. Therefore, the authors focused on yearly lagged values of the key variables to analyze their causal relationships. Moreover, this specification of the model fits well with the Granger causality.

Desjardins et al. (2022) analyzed the causality between reinsurance demand and liquidity creation by applying a robust GMM procedure to estimate their parameters. More precisely, they used the regularized GMM procedure proposed by Carrasco and Nayihouba (2022) for dynamic panel data. They estimated equation (48) where $y_{i,t}$ is for reinsurance demand and $x_{i,t}$ is for the liquidity creation ratio:

$$\begin{cases} y_{i,t} = \beta_1 x_{i,t-1} + \beta_2 y_{i,t-1} + \Delta_1 Z_{i,t} + \kappa_i + \varepsilon_{i,t} \\ x_{i,t} = \beta_3 x_{i,t-1} + \beta_4 y_{i,t-1} + \Delta_2 \Xi_{it} + \theta_i + \nu_{i,t} \end{cases} \tag{48}$$

In equation (48), the liquidity creation ratio at time *t* is regressed on the control variables at time *t,* and the reinsurance demand at time *t* is regressed on the control variables at time *t*. Each equation of the model is in fact a dynamic panel data relationship with a lagged dependent variable, a lagged endogenous variable, individual fixed effects (κ_i, θ_i) and vectors of covariates $(Z_{it}, \Xi_{i,t})$. The terms $\varepsilon_{i,t}$ and $\nu_{i,t}$ are error terms with zero mean and positive variance for $i = 1...N$ and $t = 1...T$, where *N* is the number of firms, and *T* the number of periods.

The estimation of equation (48) is done equation by equation. Using the above specification of the model, we may face some severe endogeneity problems that need to be solved in the estimation process. The first endogeneity problem is due to the presence of individual fixed effects in the model, which creates a correlation between the error term and one of the explanatory variables, namely the lagged value of the dependent variable. This implies that the lagged value of the dependent variable should be treated as an endogenous variable in the estimation process. Moreover, this problem is amplified by the fact that the lagged value of liquidity creation in the equation of reinsurance demand and the lagged value of the reinsurance demand in the equation of liquidity creation are also endogenous variables, as was shown empirically by the authors using Hausman's (1978) test procedure. Therefore, the standard OLS method with fixed effects may yield bias estimates. To tackle these two endogeneity problems and to avoid the problem involved in finding valid instruments for the two-stage least square (2SLS) regression method, the GMM model is employed to estimate parameters in (48), with lagged levels of the set of explanatory variables as instruments. This model has an important feature: If a variable at a certain period can be used as an instrument, then all the past realizations of that variable can also be used as instruments. Therefore, the number of moment conditions can be very large, even if the time duration of panel T is finite.

The presence of this large set of moment conditions may create a variance bias that is also referred to as the many instruments problem. Moreover, the lagged levels of the dependent variable, which appear in the explanatory variables, can become weak instruments when the autoregressive parameter is close to unity (Blundell and Bond, 1998). As a solution to these problems (many instruments and weak instruments), one can add some regularization methods to the standard GMM method, as Carrasco and Nayihouba (2022) did, to evaluate the relationship between liquidity creation and reinsurance demand.

7.2 Regularization procedures for estimation

Several methods have been proposed in the context of cross-sectional data models to deal with these problems of instruments. To manage this problem in a dynamic setting, Okui (2009) recommends choosing the optimal number of moment conditions to minimize the mean square error of the estimation in order to improve the finite sample properties. However, the finite sample problem is not completely solved since there may be a large bias in estimated cross-lagged parameters when the autoregressive coefficient in the dynamic panel is close to unity. Carrasco and Nayihouba (2022) proposed a more general method based on different ways of inverting the covariance matrix of instruments. They showed that this method improves the properties of the GMM estimation even if the autoregressive coefficient is close to unity. To analyze the causality relationships in (48), Desjardins et al. (2022) focused on two of the regularization procedures proposed by Carrasco and Nayihouba (2022) in the context of their dynamic panel data.

When the number of moment conditions exceeds the number of unknown parameters to be estimated by GMM, the model validity must be confirmed, by testing the overidentifying restriction, before making any inference regarding the resulting estimation. A common test for this purpose is the J-test proposed by Sargan (1958) and Hansen (1982). To test if their models were well specified, Desjardins et al. (2022) applied the modified version of the J-test to the context of dynamic panel data models (Arellano and Bond, 1991). They found a causal relationship in the two equations, meaning that insurers who invest more in the economy buy more reinsurance for protection, and those who buy more reinsurance invest more in the economy because they are well protected.

8. Conclusion

We have explored the empirical measurement of asymmetric information on resource allocation. We have studied causality effects with applied DID methodology (with parallel trends or propensity score modeling) and with instrumental variables (with essential heterogeneity or not). Other tests were applied with dynamic data. We did not cover the literature with dynamic treated effects because our applications did not contain such effects.

Three information problems drew our attention: moral hazard, asymmetric learning, and adverse selection. One conclusion that seems to be accepted by many authors is that information problems may create distortions in the economy, in contrast with a situation of full and perfect information. Effective mechanisms have been established to reduce these distortions and to eliminate residual information effects at the margin. In this new version of our survey, we have emphasized the role of causality to identify different information problems. We have shown that causality, in dynamic models, can be used to separate unobserved heterogeneity from moral hazard and to apply tests to separate moral hazard from adverse selection and asymmetric learning. We have also shown how instrumental variable and DID methodologies permit to identify moral hazard in auction and insurance applications, and to identify the effect of a new regulation in the CDS market. Finally, we have used the GMM methodology to study the dynamic links between reinsurance demand and liquidity creation.

We have studied the causal effect of risk management on firm value and risk. We have shown that when appropriate methodology is applied (instrumental variable with essential heterogeneity), risk management is a causal source of improvement in firm value and risk for non-financial firms.

References

Abadie A (2003) Semiparametric instrumental variable estimation of treatment response model. Journal of Econometrics 113:231-263

Abadie A (2021) Using synthetic controls: Feasibility, data requirements, and methodological aspects. Journal of Economic Literature 59(2):391-425

Abadie A, Diamond A, Hainmueller J (2010) Synchetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. Journal of the American Statistical Association 105(490):493-505

Abadie A, Diamond A, Hainmueller J (2015) Comparative politics and the synthetic control method. American Journal of Political Science 59(2):495-510

Abadie A, Gardeazabal J (2003) The economic costs of conflict: A case study of the Basque country. American Economic Review 93(1):113-132

Abadie A, L'Hour J (2021) A penalized synthetic control estimator for disaggregated data. Journal of the American Statistical Association 116(536):1817-1834

Abay K (2018) How effective are non-monetary instruments for safe driving? Panel data evidence on the effect of the demerit point system in Denmark. Scandinavian Journal of Economics 120(3):894-924

Abbring JH, Chiappori PA, Pinquet J (2003) Moral hazard and dynamic insurance data. Journal of the European Economic Association 1:767-820

Adam T, Fernando, CS (2006) Hedging, speculation, and shareholder value. Journal of Financial Economics 81:283-309

Akari MA, Ben-Abdallah R, Breton, M, Dionne, G (2021) The impact of central clearing on the market for single-name credit default swaps. North American Journal of Economics and Finance 56(101346)

Akerlof GA (1970) The market for 'lemons': Quality uncertainty and the market mechanism. Quarterly Journal of Economics 84:488-500

Allayannis G, Weston JP (2001) The use of foreign currency derivatives and firm market value. Review of Financial Studies 14:243–276

Angrist JD (1990) Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records. The American Economic Review 80(3):313-336

Angrist JD (2022) Empirical strategies in economics: Illuminating the path from cause to effect. Econometrica 90(6):2509-2530

Angrist JD, Graddy K, Imbens GW (2000) The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. The Review of Economic Studies 67(3):499-527

Angrist JD, Hahn J (2004) When to control for covariates? Panel asymptotics for estimates of treatment effects. The Review of Economics and Statistics 86(1):58-72

Angrist JD, Imbens GW (1995) Two-stage least squares estimation of average causal effects in models with variable treatment intensity. Journal of the American Statistical Association 90(430):431-442

Angrist JD, Krueger A (1991) Does compulsory schooling affect schooling and earnings? Quarterly Journal of Economics CVI:979-1014

Arellano M, 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(2):277-297

Arkhangelsky D, Athey S, Hirshberg DA, Imbens GW, Wager S (2021) Synthetic difference-in-differences. American Economic Review 111(12):4088-4118

Arrow K (1963) Uncertainty and the welfare economics of medical care. American Economic Review 53:941-973

Ashenfelter O, Card D (1985) Using the longitudinal structure of earnings to estimate the effect of training programs. The Review of Economics and Statistics 67(4):648-660

Ashenfelter O, Heckman J (1974) The estimation of income and substitution effects in a model of family labor supply. Econometrica 42(1):73-85

Ashenfelter O, Krueger A (1994) Estimates of the economic return to schooling from a new sample of twins. The American Economic Review 84:1157-1173

Bartram SM, Brown GW, Conrad J (2011) The effects of derivatives on firm risk and value. Journal of Financial and Quantitative Analysis 46:967-999

Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? Quarterly Journal of Economics 119(1):249-275

Bloom HS (1984) Accounting for no-shows in experimental evaluation design. Evaluation Review 8:225-246

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

Blundell RW, Smith RJ (1989) Estimation in a class of simultaneous equation limited dependent variable models. The Review of Economic Studies 56(1):37-47

Bourgeon JM, Picard P (2007) Point-record driver's license and road safety: An economic approach. Journal of Public Economics 91:235-258

Boyer M, Dionne G (1989) An empirical analysis of moral hazard and experience rating. Review of Economics and Statistics LXXXI(1):128-134

Campello M, Lin C, Ma Y, Zou H (2011) The real and financial implications of corporate hedging. Journal of Finance 66:1615-1647

Card D (1990) The impact of the Mariel boatlift on the Miami labor market. Industrial and Labor Relation 43:245-257

Card D, Krueger A (1994) Minimum wages and employment: Case study of the fast-food industry in New Jersey and Pennsylvania. American Economic Review 84:772-793

Card, D (2022) Design-based research in empirical economics. American Economic Review 112:1773-1781

Carrasco M, Nayihouba A (2022) Regularized estimation of dynamic panel models. First view article, Econometric Theory

Carter DA, Rogers D, Simkins BJ (2006) Does hedging affect firm value? Evidence from the US airline industry. Financial Management 35:53-87

Chiappori PA (1994) Théorie des contrats et économétrie de l'assurance: quelques pistes de recherche, Mimeo, Chaire d'économie et d'économétrie de l'assurance, EHESS - ENSAE, DELTA

Chiappori PA (2000) Econometric models of insurance under asymmetric information. In: Dionne G (ed) Handbook of Insurance, Springer, p 365-393

Chiappori PA, Macho I, Rey P, Salanié B (1994). Repeated moral hazard: The role of memory, commitment, and the access to credit markets. European Economic Review 38:1527-1553

Chiappori PA, Salanié B (2000) Testing for asymmetric information in insurance markets. Journal of Political Economy 108:56-78

Chiappori PA, Salanié B (2013) Asymmetric information in insurance markets: Predictions and tests. In: Dionne G. (ed) Handbook of Insurance 2nd edn. Springer, New York, p 397-422

Cohen A (2005) Asymmetric information and learning in the automobile insurance market. Review of Economics and Statistics 87:197-207

Cohen A, Einab L (2007) Estimating risk preferences from deductible choices. American Economic Review 97(3):745-788

Cohen A, Siegelman P (2010) Testing for adverse selection in insurance markets. Journal of Risk and Insurance 77(1):39-84

Cox DR (1972) Regression models and life tables. Journal of the Royal Statistical Society, Series B 34:187-220

Crocker K, Snow A (1985) The efficiency of competitive equilibria in insurance markets with asymmetric information. Journal of Public Economics 26:207-219

Crocker K, Snow A (1986) The efficiency effect of categorical discrimination in the insurance industry. Journal of Political Economy 94:321-344

Crocker K, Snow A, Rothschild (2024) The theory of risk classification. In: Dionne G (ed) Handbook of Insurance 3rd edn. Springer, New York

Currie J, Kleven H, Zwiers E (2020) Technology and big data are changing economics: Mining text to track methods. AEA Papers and Proceedings 110:42-48

D'Arcy SP, Doherty NA (1990) Adverse selection, private information, and lowballing in insurance markets. The Journal of Business 63(2):145-64

Dahlby BG (1983) Adverse selection and statistical discrimination. An analysis of canadian automobile insurance. Journal of Public Economics 20:121-130. Reprinted in: Dionne G, Harrington S (eds) (1992) Foundations of insurance economics - Readings in economics and finance. Kluwer Academic Publishers

Dahlby BG (1992). Testing for assymmetric information in Canadian automobile insurance. In: Dionne G (ed) Contributions to insurance economics. Kluwer Academic Publishers

Dardanoni V, Forcina A, Li Donni P (2018) Testing for asymmetric information in insurance markets: A multivariate ordered regression approach. Journal of Risk and Insurance 85(1):107-125

De Meza D, Webb DC (2001) Advantageous selection in insurance markets. Rand Journal of Economics 32(2):249-262

Dehejia RH, Wahba S (1999) Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. Journal of the American Statistical Association 94:1053-1062

Desjardins D, Dionne G, Koné N (2022) Reinsurance demand and liquidity creation: A search for bicausality. Journal of Empirical Finance 66:137-154

Devlin RA (2002) Determinants of no-fault insurance measures. Journal of Risk and Insurance 69(4):555-576

Dionne G (2001) Insurance regulation in other industrial countries. In: Cummins JD (ed) Deregulating property–liability insurance. AEI–Brookings, Washington, p 362-390

Dionne G (2019) Corporate Risk Management: Theories and Applications. Wiley

Dionne G, Doherty N (1994) Adverse selection, commitment and renegotiation: Extension to and evidence from insurance markets. Journal of Political Economy 102(2):209-235. Reproduced in: Niehaus GR (ed) Insurance and Risk Management, vol. I, Economics of Insurance Markets. Elgar Reference Collection, USA

Dionne G, Fenou A, Mnasri M (2023) Insurers' M&A in the United States during the 1990-2022 period. Mimeo, Canada Research Chair in Risk Management, HEC Montréal

Dionne G, Fombaron N, Mimra W (2024) Adverse selection in insurance. In: Dionne G (ed) Handbook of Insurance 3rd edn. Springer, New York

Dionne G, Gagné R (2001) Deductible contracts against fraudulent claims: Evidence from automobile insurance. Review of Economics and Statistics 83(2):290-301

Dionne G, Garand M (2003) Risk management determinants affecting firms' values in the gold mining industry: New empirical results. Economic Letters 79:43-52

Dionne G, Ghali O (2005) The (1992) bonus-malus system in Tunisia: An empirical evaluation. Journal of Risk and Insurance 72:609-633

Dionne G, Gouriéroux C, Vanasse C (2001) Testing for evidence of adverse selection in the automobile insurance market: A comment. Journal of Political Economy 109:444-453

Dionne G, Gouriéroux C, Vanasse C (2006) The informational content of household decisions with applications to insurance under asymmetric information. In Chiappori PA and Gollier C (eds) Competitive failures in insurance markets, MIT Press Book, p 159-184

Dionne G, Gueyie JP, Mnasri M (2018) Dynamic corporate risk management: Motivations and real implications. Journal of Banking & Finance 95:97-111

Dionne G, Harrington SE (2014) Insurance and insurance markets. In: Viscusi WK and Machina M (eds) Handbook of the Economics of Risk and Uncertainty, 1st edn. North Holland, Amsterdam, p 203-261

Dionne G, La Haye M, Bergerès AS (2015) Does asymmetric information affect the premium in mergers and acquisitions? Canadian Journal of Economics 48(3):819-852

Dionne G, Lasserre P (1985) Adverse selection, repeated insurance contracts and announcement strategy. Review of Economic Studies 70(4):719-724

Dionne G, Lasserre P (1987) Adverse selection and finite–horizon insurance contracts. European Economic Review 31(4):843-862

Dionne G, Liu Y (2021) Effects of insurance incentives on road safety: Evidence from a natural experiment in China. Scandinavian Journal of Economics 123(2):453-477

Dionne G, Maalaoui Chun O, Triki T (2019) The governance of risk management: The importance of directors' independence and financial knowledge. Risk Management and Insurance Review 22:247-277

Dionne G, Michaud PC, Dahchour M (2013a) Separating moral hazard from adverse selection and learning in automobile insurance: Longitudinal evidence from France. Journal of the European Economic Association 11:897-917

Dionne G, Michaud PC, Pinquet J (2013b) A review of recent theoretical and empirical analyses of asymmetric information in road safety and automobile insurance. Research in Transportation Economics 43:85-97

Dionne G, Mnasri M (2018) Real implications of corporate risk management: Review of main results and new evidence from a different methodology. L'Actualité économique 94(4):407-452

Dionne G, Ouederni K (2011) Corporate risk management and dividend signaling theory. Finance Research Letters 8:188-195

Dionne G, Pinquet J, Maurice M, Vanasse C (2011) Incentive mechanisms for safe driving: A comparative analysis with dynamic data. Review of Economics and Statistics 93:218-227

Dionne G, Rothschild C (2014) Economic effects of risk classification bans. The Geneva Risk and Insurance Review 39:184-221

Dionne G, St-Amour P, Vencatachellum D (2009) Asymmetric information and adverse selection in Mauritian slave auctions. Review of Economic Studies 76:1269-1295

Dionne G, Triki T (2013) On risk management determinants: What really matters? European Journal of Finance 19:145-164

Dionne G, Vanasse C (1989) A generalization of automobile insurance rating models: The negative binomial distribution with a regression component. Astin Bulletin 19(2):199-212

Dionne G, Vanasse C (1992) Automobile insurance ratemaking in the presence of asymmetrical information. Journal of Applied Econometrics 7:149-165

Fang H, Keane MP, Silverman D (2008) Sources of advantageous selection: Evidence from the Medigap insurance market. Journal of Political Economy 116(2):303-350

Finkelstein A, McGarry K (2006) Multiple dimensions of private information: Evidence from the long-term care insurance market. American Economic Review 96:938-958

Frisch R (1930) A dynamic approach to economic theory: Lectures by Ragnar Frisch at Yale University. Lectures at Yale University beginning September 1930. Mimeographed, Frisch Archives, Department of Economics, University of Oslo

Frisch R (1938) Autonomy of economic relations: Statistical versus theoretical relations in economic macrodynamics. Paper given at League of Nations. Reprinted in: Hendry DF and Morgan MS (1995). The Foundations of Econometric Analysis, Cambridge University Press.

Froot KA, Scharfstein D, Stein J (1993) Risk management: Coordinating corporate investment and financing policies. Journal of Finance 48:1629-1658

Gale D, Hellwig M (1985) Incentive-compatible debt contracts: The one-period problem. The Review of Economic Studies 5(4):647-663

Gay GD, Lin CM, Smith SD (2011) Corporate derivatives use and the cost of equity. Journal of Banking & Finance 35:1491-1506

Geyer A, Kremslehner D, Mürmann A (2020) Asymmetric information in automobile insurance: Evidence from driving behavior. Journal of Risk and Insurance 87(4):969-995

Gormley TA, Matsa DA (2011) Growing out of trouble? Corporate responses to liability risk. The Review of Financial Studies 24(8):2781-2821

Gouriéroux C (1999) The econometrics of risk classification in insurance. Geneva Papers on Risk and Insurance Theory 24:119-138

Goux D, Maurin E (2023) David Card, Nobel Prize 2021: The Design-Based Revolution. Revue d’économie politique 133(1):25-46

Graham JR, Rogers DA (2002) Do firms hedge in response to tax incentives? Journal of Finance 57(2):815-839

Graham JR, Smith CW (1999) Tax incentives to hedge. Journal of Finance 54:2241-2262

Guay WR (1999) The impact of derivatives on firm risk: An empirical examination of new derivatives users. Journal of Accounting and Economics 26:319-351

Haavelmo T (1943) The statistical implications of a system of simultaneous equations. Econometrica 11(1):1-12

Haavelmo T (1944) The probability approach in econometrics. Econometrica 12 (supplement) iii-vi:1-115

Hansen LP (1982) Large sample properties of generalized method of moments estimators. Econometrica 50(4):1029-1054

Haushalter D (2000) Financing policy, basis risk, and corporate hedging: Evidence from oil and gas producers. Journal of Finance 55:107-152

Hausman JA (1978) Specification tests in econometrics. Econometrica 46:1251-1272

Heckman JJ (1990) Varieties of selection bias. The American Economic Review 80:313-318

Heckman JJ, Humphries JE, Veramendi G (2016) Dynamic treatment effects. Journal of Econometrics 191:276-292

Heckman JJ, Urzua S, Vytlacil EJ (2006) Understanding instrumental variables in models with essential heterogeneity. Review of Economics and Statistics 88:389-432

Heckman JJ, Vytlacil EJ (1999) Local instrumental variables and latent variable models for identifying and bounding treatment effects. Proceedings of the National Academy of Sciences 96(8): 4730-4734

Heckman JJ, Vytlacil EJ (2001) Policy-relevant treatment effects. Papers and proceedings of the 113th annual meeting of the American Economic Association 91(2):107-111

Heckman JJ, Vytlacil EJ (2005) Structural equations, treatment effects and econometric policy evaluation. Econometrica 73(3):669-738

Heckman JJ, Pinto R (2022) Causality in econometrics. NBER Working paper 29787.

Hendel I, Lizzeri A (2003) The role of commitment in dynamic contracts: Evidence from life insurance. The Quarterly Journal of Economics 118(1):299-327

Hendry DF (1980) Econometrics – alchemy or science? Economica 47(188):387-406

Hirano K, Imbens GW., Ridder G (2003) Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71(4):1161–1189

Holland PW (1986) Statistics and causal inference. Journal of the American Statistical Association 81:945-960

Holmström B (1979) Moral hazard and observability. The Bell Journal of Economics 10(1):74-91

Hoy M (1982) Categorizing risk in the insurance industry. Quarterly Journal of Economics 97:321-336

Hoyt RE, Liebenberg AP (2024) Enterprise risk management. Handbook of Insurance 3rd edn. Springer, New York

Imbens GW (2022) Causality in econometrics: Choice vs chances. Econometrica 90(6):2541-2566

Imbens GW, Angrist JD (1994) Identification and estimation of local average treatment effects. Econometrica 61:467-476

Imbens GW, Rubin DB (2015) Causal inference in statistics, social, and biomedical sciences. New York, Cambridge University Press

Imbens GW, Wooldridge JM (2009) Recent developments in the econometrics of program evaluation. Journal of Economic Literature 47(1):5-86

Jin Y, Jorion P (2006) Firm value and hedging: Evidence from U.S. oil and gas producers. Journal of Finance 61:893-919

Kang JK, Kim JM (2008) The geography of block acquisitions. The Journal of Finance 63(6): 2817-2858

Kilian L (2009) Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. American Economic Review 99:1053-1069

Kim H, Kim D, Im S, Hardin JW (2009) Evidence of asymmetric information in the automobile insurance market: Dichotomous versus multinomial measurement of insurance coverage. Journal of Risk and Insurance 76(2):343-366

Lalonde RJ (1986) Evaluating the econometric evaluation of training programs with experimental data. The American Economic Review 76(4):604-620

Leamer EE (1978) Specification searches: Ad hoc inference with nonexperimental data. Wiley Series in Probability and Mathematical Statistics

Leamer EE (1983) Let's take the con out of econometrics. The American Economic Review 73(1):31-43

Lee DS, Lemieux T (2010) Regression discontinuity designs in economics. Journal of Economic Literature 48:281-355

Lee YW (2013) Testing for the presence of moral hazard using the regulatory reform in the car insurance market: Case of Korea. Japanese Economic Review 64:414-429

Low H, Meghir C (2017) The use of structural models in econometrics. Journal of Economic Perspectives 31(2):33-58

Maliar S, Salanié B (2022) Testing for asymmetric information in insurance with deep learning. Working paper, Columbia University, 31 p.

Manski CF (1990) Nonparametric bounds on treatment effects. The American Economic Review 80:319-323

Mayers D, Smith CW (1982) On the corporate demand for insurance. The Journal of Business 55:281-296

Mincer J (1974) Schooling, experience and earnings. New York: National Bureau of Economic Research

Mnasri M, Dionne G, Gueyie JP (2017) The use of nonlinear hedging strategies by US oil producers: Motivations and implications. Energy Economics 63:348-364

Neyman J (1923) Statistical problems in agricultural experiments. Journal of the Royal Statistical Society II (Supplement)2:107-180

Okui R (2009) The optimal choice of moments in dynamic panel data models. Journal of Econometrics 151:1-16

Olivella P, Vera-Hernández M (2013) Testing for asymmetric information in private health insurance. The Economic Journal 123(567):96-130

Parra A, Winter R (2024) Optimal insurance contracts under moral hazard. In: Dionne G (ed) Handbook of Insurance 3rd edn. Springer, New York

Pauly MV (1968) The economics of moral hazard: Comment. The American Economic Review58(3):531-537

Pearl J (2009) Causal inference in statistics: An overview. Statistics Surveys 3:96-146

Pearl J (2012) The do-calculus revisited. CORR abs/1210.4852

Pérez-Gonzalez F, Yun H (2013) Risk management and firm value: Evidence from weather derivatives. Journal of Finance 68:2143-2176

Picard P (2024) Economic analysis of insurance fraud. In: Dionne G (ed) Handbook of Insurance 3rd edn. Springer, New York

Pinquet J (2024) Nonnegative second-order semiparametric analysis and experience rating in non-life insurance. In: Dionne G (ed) Handbook of Insurance 3rd edn. Springer, New York

Puelz R, Snow A (1994) Evidence of adverse selection: equilibrium signaling and cross-subsidization in the insurance market. Journal of Political Economy 102:236-257

Richaudeau D (1999) Automobile insurance contracts and risk of accident: An empirical test using french individual data. Geneva Papers on Risk and Insurance Theory 24(1):97-114

Rosenbaum PR, Rubin DB (1983a) Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. Journal of the Royal Statistical Society Series B (Methodological):212-218

Rosenbaum PR, Rubin DB (1983b) The central role of the propensity score in observational studies for causal effects. Biometrika 70:41-55

Rothschild M, Stiglitz J (1976) Equilibrium in competitive insurance markets : An essay on the economics of imperfect information. Quarterly Journal of Economics 90:629-650. Reprinted in: Dionne G, Harrington S (eds) (1992) Foundations of insurance economics - Readings in economics and finance. Kluwer Academic Publishers

Rowell D, Nghiem SH, Connelly, LB (2017a) Two tests for ex ante moral hazard in a market for automobile insurance. Journal of Risk and Insurance 84:1103-1126

Rowell D, Nghiem SH, Connelly LB (2017b) Testing for asymmetric information in insurance markets: A test for ex ante moral hazard revisited. Economics Letters 150:4-5

Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 66(5):688-701

Saito K (2006) Testing for asymmetric information in the automobile insurance market under rate regulation. Journal of Risk and Insurance 73(2):335-356

Sargan JD (1958) The estimation of economic relationships using instrumental variables. Econometrica 26:393-415

Shavell S (1979) Risk sharing and incentives in the principal and agent relationship. The Bell Journal of Economics 10(1):55-73

Smith CW, Stulz RM (1985) The determinants of firms' hedging policies. Journal of Financial and Quantitative Analysis 20(4):391-405

Spindler M (2014) Econometric methods for testing for asymmetric information: A comparison of parametric and nonparametric methods with an application to hospital daily benefits. The Geneva Risk and Insurance Review 39:254–266

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

Stulz RM (1984) Optimal hedging policies. Journal of Financial and Quantitative Analysis 19:127-140

Stulz RM (1996) Rethinking risk management. Journal of Applied Corporate Finance 9:8-24

Su L, Spindler M (2013). Nonparametric testing for asymmetric information. Journal of Business & Economic Statistics 31(2):208-225

Sun L, Abraham S (2021) Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics 225:175-199

Thistlethwaite DL, Campbell DT (1960) Regression discontinuity analysis: An alternative to the ex post facto experiment, Journal of Educational Psychology 51(6):309-317

Townsend RM (1979) Optimal contracts and competitive markets with costly state verification. Journal of Economic Theory 21(2):265-293

Tufano P (1996) Who manages risk? An empirical examination of risk management practices in the gold mining industry. Journal of Finance 51:1097-1137

Wilson C (1977) A model of insurance markets with incomplete information. Journal of Economic Theory 16:167-207

Wooldridge JM (2010) Econometric Analysis of Cross Section and Panel Data, 2nd ed. MIT Press, Boston

Zavadil T (2015) Do the better insured cause more damage? Testing for asymmetric information in car insurance. Journal of Risk and Insurance 82:865-889