Financial History Review 22.1 (2015), pp. 1–18. © European Association for Banking and Financial History e.V. 2015
doi:10.1017/S0968565015000025

THE PAST MIRROR: NOTES, SURVEYS, DEBATES

Identifying economic crises: insights from history

XAVIER DE SCHEEMAEKERE, KIM OOSTERLINCK and ARIANE SZAFARZ
Université Libre de Bruxelles

Economists have been blamed for their inability to forecast and address crises. This article attributes this inability to intertwined factors: the lack of a coherent definition of crises, the reference-class problem, the lack of imagination regarding the nature of future crises and sample-selection biases. Specifically, economists tend to adapt their views on crises to recent episodes, and omit averted and potential crises. Threshold-based definitions of crises run the risk of being ad hoc. Using historical examples, this article highlights some epistemological shortcomings of the current approach.

Keywords: economic crisis, economic history, single-case probability, epistemology

JEL classification: B40, G01, G17, N00

Imagination is more important than knowledge. For knowledge is limited, whereas imagination embraces the entire world, stimulating progress, giving birth to evolution.
Albert Einstein[1]

I

In a now famous visit to the London School of Economics in November 2008, Queen Elizabeth II asked why nobody had noticed the turmoil on the international markets before the crisis occurred (Pierce 2008). According to the *Huffington Post*,

X. De Scheemaekere, Université Libre de Bruxelles, SBS-EM, CEB; K. Oosterlinck, corresponding author: Université Libre de Bruxelles, SBS-EM, CEB, kooster l@ulb.ac.be; Ariane Szafarz, Université Libre de Bruxelles, SBS-EM, CEB and CERMi; Av. F. D. Roosevelt, 50, 1050 Brussels, Belgium. The authors would like to thank Marie Brière, two anonymous referees as well as the editor for helpful suggestions and comments.

[1] As quoted in 'What life means to Einstein: an interview by George Sylvester Viereck' in the *Saturday Evening Post*, 26 Oct. 1929.

following this question a group of eminent economists apologized to the Queen and blamed 'a failure of the collective imagination of many bright people … to understand the risks to the system as a whole'. Nevertheless, the economists' failure to forecast and address the recent crisis is more than an isolated event. Similar drawbacks have been highlighted in the past as well. Davies and McGoey (2012) argue that ignorance is sometimes a convenient excuse for the absence of reaction to potential crises, and ultimately for the denial of responsibility. Undoubtedly, lack of forward-looking imagination partly explains why economists systematically miss the outbreak of crises. This article, however, argues that lack of imagination is but one manifestation of a more general problem ingrained in the way economists identify crises.

The standard approach in economics uses statistical techniques to understand the world.[2] In particular, economists concerned with crises typically rely on statistics to infer crisis probabilities and draft policy responses. The legitimacy of such a frequentist approach, however, depends on specific assumptions. First, the object under study – here, the crisis – is well defined. Second, any observed crisis is attributed to the appropriate reference class (as crises of different natures exist). Third, the sample used to infer probabilities is not biased. This article argues that none of these assumptions is actually met in practice. Let us develop our point with a metaphor.

Consider two friends, Jack and Jill (J&J), who frequently play games. They have played chess ten times and from the evidence we have been able to gather, we estimate that Jill won eight times and Jack twice. And they have played tennis ten times, with estimates giving Jill four wins and Jack six. Today J&J are playing a new board game whose rules have never been tested before. What is the probability of Jack winning? This seemingly simple question raises issues regarding definition, reference class and sample bias.

First, the issues pertaining to definition are the following. What counts as a win or as a success? Perhaps we have misclassified the past events, because we only see the physical movements of J&J and do not know the winning conditions. Is the new game unique and are all past elements therefore irrelevant? What do we know about the next game J&J will play?

Second, we have the reference-class problem. The computed probabilities will depend on the chosen reference class. One could argue that the probability of Jack winning the new game is 0.4, since Jack wins 40 per cent of games. Alternatively, one could say that the probability is 0.2, because Jack wins 20 per cent of board games. Eventually, one could also consider the probability as undefined since J&J have never played the new board game.

Last, there is an issue associated with sample biases. Do we really know that there have been ten chess games and ten tennis games? Perhaps we really know that Jack has won two chess games and six tennis games, but do not have a great deal of evidence about how many games J&J have played. Did J&J report the results correctly? Perhaps

[2] According to Morgan (1990), econometrics became the dominant form of applied economics after 1940. The importance of statistics in economic thinking has been such that some subfields of statistics owe their development to financial markets.

https://doi.org/10.1017/S0968565015000025 Published online by Cambridge University Press

they forgot to report unfinished games or games that ended in a tie. Should we just use the frequencies so far of Jack winning, or the frequencies of Jack winning in the past and in the future? Should we exclude the chess game for which Jack reported having a headache? How do we imagine future games? It might be the case that the new game is so different from the previous ones that past games do not inform us on winning likelihood. In this case, alternative forms of reasoning, imagination in a sense, may prove more fruitful than a frequentist approach.

These three types of problem are mostly ignored by economists. They are in fact inextricably intertwined. Typically, existing studies fail to define crises unambiguously. The classifications found in the literature change through time and are inconsistent with each other. As a result statistical analyses are likely to suffer from a reference-class problem. Ambiguous definitions and badly defined reference classes increase the likelihood that studies will rely on biased samples. Eventually, using past-frequentist approaches is more likely to lead to a failure to imagine new types of crises. We contend that only a more conceptual approach to crises can overcome these limitations. To clarify its arguments, this article builds heavily on historical examples.

The rest of the article is organized as follows. The next section discusses how economists identify and classify crises. Section III shows how the reference-class problem may lead to biased conclusions. Section IV presents the biases stemming from the ambiguity regarding crisis definitions and the reference-class problem. Section V concludes.

II

As in J&J's case, the ability to define correctly what represents a success or a failure is crucial for determining probabilities of occurrence of a given event. Regarding economic crises one may first ask: what are economic crises? How should we classify them? Economists are mostly unable to answer these questions. Instead, they tend to subjectively adapt their views on crises to recent episodes. As pointed out by Lo (2012), there is no agreement on the basic facts related to the latest crisis, let alone on its starting date. Empirical evidence is fuzzy and interpretations as well as conclusions exhibit significant heterogeneity. The problem of definition is compounded by the fact that crises shape economic discourses (Flandreau 2012). This in turn affects the generally agreed-upon narratives on specific crises and therefore the way they are perceived. Classifications of economic crises come in many forms. Some authors track 'abnormal' data, and declare that above or below a certain threshold there is a crisis. For instance, for Reinhart and Rogoff (2004) a free-falling episode is defined by a 12-month inflation threshold of 40 per cent or more. For Reinhart and Rogoff (2011), an inflation crisis is an episode during which inflation exceeds 20 per cent per year, while hyperinflation means inflation above 500 per cent per year.

Defining crises by means of thresholds and outliers has intuitive appeal. It gives the impression that crises share certain features. This is, however, valid if and only if the underlying variables are driven by an objective and invariant probability distribution. It may be wondered to what extent all so-called hyperinflations do indeed stem from a

https://doi.org/10.1017/S0968565015000025 Published online by Cambridge University Press

unique distribution. Is there any reason to believe that the 1946 Hungarian hyperinflation was generated by the same underlying distribution as Zimbabwe's hyperinflation in 2008? In some instances, the underlying structural relationship is obviously unstable. For example, Barkbu *et al.* (2012) stress the changes in both crises and the IMF programmes over time. Alternatively, one could argue that each relevant hyperinflation episode is the result of a conjunction of economic circumstances driven by country-specific factors. Some Latin American countries have long lived with double-digit inflation rates, while European countries would view such a situation as a disaster. Threshold-based definitions of crises run the risk of being ad hoc. They lack theoretical grounds. For instance, data-driven thresholds for delineating hyper-inflationary episodes are arbitrary.[3] Why would crises start at 20 per cent inflation? Why not 19 per cent or 21 per cent? Although robustness checks can at least partially address this issue, threshold-based definitions of crises are both inelegant and unconvincing.

Most importantly, the threshold-based approach inevitably neglects crises averted just before their outbreak. Looking only at symptoms is insufficient to provide preventive cures. In medicine such an approach would rule out research aimed at understanding why some patients contract a disease while others remain immune. In the hyperinflation example, scrutinizing countries that managed to combat a borderline-high inflation rate could be as insightful as restricting the analysis to those who fell into the trap.

The standard empirical approach to crises is based on statistical thresholds rationalized by the supposed 'anomalies' associated with them. From a probabilistic standpoint, this hardly makes sense because extreme events are naturally consistent with low-probability outcomes. This is the motivation for using, say, 5 per cent thresholds in hypothesis testing. If economic crises were merely below- or above-threshold outcomes driven by otherwise-standard theory, their representation would require no special attention. In contrast, what economists have in mind when speaking about crises are most likely specific features that make the underlying model behave differently than during 'normal' times. Therefore, threshold-based econometrics is inappropriate for dealing with crises. Rather, if crises were special, then asymptotic theory based on standard models would be discarded.[4] As a result, econometrics alone is unable to properly address and classify crises. There is a strong need for a theory that does not rely on past observations but rather on an ex ante conceptual approach.

A refinement of the threshold-based approach consists in combining it with expert knowledge. In this line of thought, Schularick and Taylor (2012) use a two-step process to identify banking crises. First, they select 'events during which a country's

[3] Hyperinflation is just an illustration. Setting a threshold to define a crisis is probably the most common approach to economic crises.

[4] Moreover, crises are supposedly rare and would then create small-sample distortions. As well as being highly sensitive to distributional assumptions, empirical results driven from small samples may exhibit large variations with respect to the inclusion/exclusion of each single observation.

https://doi.org/10.1017/S0968565015000025 Published online by Cambridge University Press

banking sector experiences bank runs, sharp increases in default rates accompanied by large losses of capital that result in public intervention, bankruptcy, or forced merger of financial institutions' (Schularick and Taylor 2012, p. 1038). Second, they send their selected dates to international experts for confirmation. Even though this refined approach has strong points, it still fails to provide an ex ante definition of crises. The assumption behind this approach is that expert knowledge is based on some obscure – and most likely incommunicable – theoretical fundaments. Indeed, if theory were clear, it would allow any economist to identify crises with no need for external confirmation.[5]

III

While a consensual definition is still lacking, many articles provide stylized facts and classify crises. As expected, classifications vary across studies because observed crises differ across a wide range of features, including the role of fundamentals, the relative importance of banks and securitized debts,[6] the relative importance of private and sovereign debts, the exchange rate regimes and history, and the underlying structure and dynamics (Portes 1999).[7] As a consequence, studies are hardly comparable. Empirical studies classify crises based on past occurrences and therefore ignore both averted and potential crises. A consensual classification should be general enough to encompass all types of imaginable crisis situations.[8] Besides the issues related to definition, researchers wishing to determine the probability that a crisis may occur should also engage in a reflection on the reference-class problem.[9]

[5] As a matter of fact, a large part of the economics profession is indifferent to this issue. The argument is, however, valid independently of how many economists actually attempt to determine probabilities for crises.

[6] Comparing the crises of 1931 and 2008, Allen and Moessner (2012 p. 123) stress that 'there was a key difference between the two crises in the range and nature of assets that were regarded as liquid and safe'.

[7] For instance, Portes (1999) and Eichengreen and Portes (1987) suggest that generalized financial crises are of three sorts: banking crises, debt crises and foreign exchange crises. Radelet and Sachs (1998) distinguish between speculative attacks on the exchange rate, financial panics (or bank runs), collapses of asset price bubbles, moral hazard crises and debt overhangs. Brière *et al.* (2012) differentiate currency crises, sovereign debt crises, equity or bond crash crises, corporate bankruptcies or loss of confidence (Enron-type crises), and severe external events (9/11-type crises).

[8] More broadly, the reference-class problem may affect any empirical study regardless of its methodology. This applies in particular to model-free economic studies. For instance, data mining has limited predictive power when it comes to economic crises (Berg and Pattillo 1999), most likely because of its atheoretical backward-looking design. While the early-warning-system methodology proposed by Kaminsky *et al.* (1998) seems to provide better results (Frankel and Saravelos 2012), it still fails to incorporate the structural breaks that characterize crises (Candelon *et al.* 2012). Despite their undeniable explorative interest, model-free studies are of little help in addressing the reference-class problem and building a meaningful classification of economic crises.

[9] Evidently, classification is a very useful tool for scientific analysis. Rather than criticizing classification *per se*, we point out the limitations inherent in the use of classifications for assessing crisis probabilities.

https://doi.org/10.1017/S0968565015000025 Published online by Cambridge University Press

The reference-class problem relates to the ex ante classification of uncertain events. In a probabilistic framework, observations are outcomes driven by an underlying generating distribution. Hence, ex ante classification is necessary for assessing theories using observations. In addition, valid classes must constitute a partition of all imaginable outcomes. Applied to economic crises, the reference-class problem involves imagining and classifying all possible scenarios for future crises.

The reference-class problem formulated by Reichenbach ([1935] 1978, p. 374) can be summarized as follows:

> If we are asked to find the probability holding for an individual future event, we must first incorporate the case in a suitable reference class. An individual thing or event may be incorporated in many reference classes, from which different probabilities will result. This ambiguity has been called the problem of the reference class.

Most importantly, the probability associated with an event 'can change depending on how it is classified' (Hájek 2007, p. 563). In the case of J&J's game the decision to refer to one class of events or the other would lead to very different outcomes. If the probabilities of winning the new game were assessed on the basis of wins in general, we would expect Jack to win 40 per cent of the new games. In contrast, if the reference class is board games, the win expectation would only be 20 per cent. Choosing a reference class may thus have serious implications in terms of occurrence assessments.

In practice, establishing the probability of an economic crisis comes down to delineating the reference classes to which such crises belong. Based on these classes, one could then try to evaluate the probabilities associated with each of them. The reference-class problem epitomizes the fact that there are always many possible classes. As Eagle (2004, p. 393) puts it:

> The obvious problem is that competing reference classes yield different probabilities, with no reference class standing out as the 'correct' one. Not only does the event seem to have no determinate unconditional probability, but there is no guide for the rational agent to assign one based on evidence, despite many attempts to provide one.

Grouping crises together is an issue that some authors deal with by considering each crisis as unique. The 'this time is different' syndrome underlined by Reinhart and Rogoff (2011) emphasizes that economic analysts tend to consider each new crisis as the sole member of a new class. This view was defended in 1913 by Wesley Mitchell, one of the founders of the National Bureau of Economic Research (Morgan 1990). Like Taleb's (2007) black swans, crises are rare and dramatic events, far removed from the repetitive events that probability statements are about, such as dice throws, roulette games, or even plane crashes. In roulette, getting a 17, for example, is rather rare (probability: 1/37), but no more exceptional than any other outcome. It is a rare event, not a dramatic one. On the contrary, plane crashes are dramatic events, but unfortunately frequent enough to infer their probability with an admissible confidence range (and subsequently make recommendations for lowering their occurrence).

But any event, be it a dice throw or a plane crash, is unique in time and place. Only abstraction makes it possible to move from single cases to a repeated idealized event. For example, we ignore space–time modifications when speaking of 'throwing the same dice ten times'. In doing so, we deliberately erase some 'insignificant conditions' in order to gather similar events and search for general principles. 'Categorical repetitivity of the world's facts is a classificatory abstraction [through which] rational knowledge begins, [i.e.] when facts are reduced into classes as symbol of conceptual categories, with the search for laws which transcends the unique event' (Scardovi 1988, p. 59).

Single cases do not make science. But to what extent is categorization legitimate? At what point do we lose the specificity of a phenomenon under study by putting it in a predefined class? In games of chance, ignoring the 'nitty-gritty causal details' (Millstein 2003, p. 1321) of concrete situations (physical forces, etc.) in favour of symmetry considerations to elaborate scientific theories seems reasonable. Moreover, concerning roulette or plane crashes, everyone knows precisely what is at stake; no one will include '18' in the 'getting a 17' event; nor will they include car accidents in the plane crash statistics.

The same is not true for economic crises as each so-called crisis is typically associated with a single-event story (Bernal *et al.* 2010). Consequently, the issue goes far beyond 'details' that can be ignored in order to study crises. The question 'How to define a crisis?' cannot be addressed without previously delineating what matters, and what does not, in the classification of crises.

In addition, the limitations of the frequentist approach are pointed out by several authors. Svetlova and Fiedler (2011) trace the critique back to the eighteenth-century philosopher David Hume, who questioned inductive methods and the common assumption that the future will reproduce the past. Knight (1921) distinguished risk from 'true' uncertainty. According to his view, risk relates to measurable uncertainty whereas true uncertainty is not quantifiable. Therefore, probabilities are meaningful only in case of risk, and not in case of true uncertainty. Building on this approach, Zeckhauser and Viscusi (1990) add ignorance to the picture, which results in a three-case classification. First, in situations characterized by risk, both the states of the world and their probabilities are known. Second, in uncertain situations, the states of the world are known but it is impossible to assign probabilities to them. Third, when ignorance prevails, the states of the world are no longer identified. In the same vein, John Maynard Keynes, George Shackle and Friedrich Hayek raised doubts about the application of frequentist probability measures to imperfect knowledge of future events (Svetlova and van Elst 2015).

In a nutshell, the frequentist approach to crisis probability is flawed.[10] This approach considers that the probability of event X is to be deduced from the relative

[10] Although the issue is rooted in problems typically associated with the frequentist approach to probability (Szafarz 1984; De Scheemaekere and Szafarz 2009), Hájek (2007) emphasizes that it can also affect the other interpretations of probability.

frequency of *X* within a reference class. It implies a sort of 'backward causation from future results to current chances' (Hájek 1997, p. 216). Such inference acts as if probability statements about *X* were counterfactually dependent on the future behaviour of *X*. But the probability of crises does not depend on whether they will actually occur in the future. For instance, crises linked to future technology, by definition, never occurred before the technology in question was invented. Therefore, this is not the right approach for estimating the probability of a crisis.

As long as crises are not properly defined within unambiguous reference classes in which inference is possible, seeking to establish the probability of a crisis is pointless. In modern economic theory, observations are considered as outcomes of a generating probability distribution. However, associating probability distributions with single events is far from obvious. Philosophers actually quarrel about the existence of such probabilities (Settle 1977; Milne 1986; Levy 2007; Bauman 2005, 2008).

Assigning a single event to two different classes can bring polar consequences. In this respect, the discussion on the nature of the 1931 German crisis provides a good example. Chang and Velasco (2001) distinguish three generations of crisis models: the first-generation model considers crises as an inescapable outcome of fiscal imbalances and fixed exchange rates. In the second-generation model, central banks may decide to abandon an exchange rate peg when defending it is too costly in terms of employment. The third-generation model stresses the role of financial institutions and the domestic banking system. Should the 1931 German crisis be analysed with the first- or third-generation model? According to Temin (2008), this difference matters not only for theoretical analysis but also for policy purposes. Indeed, if the German crisis of 1931 was just a currency crisis, the blame would rest on the government's shoulders. Alternatively, if this crisis qualifies as a twin crisis, bankers would bear the responsibility.[11] Thus, the classification of the 1931 crisis affects both economic interpretation and policy recommendations.

The role of the Empire effect on colonial borrowing costs provides another example for which choices made regarding the reference class dramatically alter the results. First, Obstfeld and Taylor (2003) and Ferguson and Schularick (2006) treat bonds issued by colonies essentially as sovereigns. To analyse the impact of colonial status on sovereign borrowing they include a dummy variable in a model aimed at explaining the spreads between sovereign bonds and the British consol. Such an approach assumes that colonial 'sovereign' bonds and bonds issued by traditional sovereigns belong to the same class of assets, namely sovereign bonds. Empirically the authors get contradictory results: in one instance being a colony would lower borrowing costs in the other case it would not. Second, according to Accominotti *et al.* (2011) colonial 'sovereign' bonds do not belong to the same reference class as traditional sovereign bonds. The colonial status embeds implicit benefit from an imperial

[11] In the same line of argument, Szafarz (2009 and 2012) discusses potential responsibilities for crises in efficient financial markets.

https://doi.org/10.1017/S0968565015000025 Published online by Cambridge University Press

guarantee. As a result, belonging to the Empire has a structural effect rather than a marginal one.

Many other examples of reference-class problems may be found in the literature. For instance, Barkbu *et al.* (2012) discuss the impact of grouping the Russian crisis of 1998 with the subsequent Latin American and Turkish crises and conclude that 'any taxonomy of crisis episodes is controversial' (Barkbu *et al.* 2012, p. 423). The Eurozone sovereign debt crisis provides an additional example of a reference-class problem. In the 1990s leading scholars such as Eaton and Fernandez (1995) considered that sovereign debt crises only concerned emerging markets as major economies seemed safe from default. This amounts to distinguishing between different classes for 'sovereign debt defaults in emerging markets' and a more general class 'sovereign debt defaults'. There is a priori no correct class even though the probabilities of default would change dramatically depending on the preferred definition.

IV

The ambiguity related to the definition of crises and the reference class may generate additional sample biases. In the case of J&J's game, how interrupted games are treated affects our ability to predict future winners. Omitting a tennis game during which Jack won four sets out of five may bias our expectations. The same argument holds for crises which were averted. In many instances economists just discard data because they consider it abnormal, but is such an approach warranted? Similarly, in J&J's game should the chess game for which Jack claimed to have had a headache be excluded? Eventually, if the new game played by J&J is so different, can we rely on our past observation to predict an outcome, or should we rely solely on our deductive analysis and imagination? Basically, are new forms of crises predictable on the basis of past data?

In the current state of economic knowledge, the onset of crises remains largely unpredictable. Admitting this fact instead of changing theory after each crisis will avoid the pervasive 'this time is different' syndrome (Reinhart and Rogoff 2011). To illustrate this point, we put forward some common distortions stemming from ad hoc analyses of crises. In particular, we pinpoint the consequences of neglecting averted crises. Such 'unborn' crises, often referred to as 'peso problems', are typically acknowledged in specific contexts, but ignored when it comes to classifying crises on a large scale. We then discuss the sample bias implied by the decision to treat some data as abnormal before discussing the role of imagination.

First, ex post views on crises prevent researchers from factoring in events that were perceived as dramatic when they occurred but failed to leave a long-lasting footprint. In particular, the so-called 'peso problem' relates to ex post implications of a disruptive event that was expected to happen, but actually did not (Veronesi 2004). As Sill (2000, p. 4) puts it: 'Peso problems can arise when the possibility that some infrequent or unprecedented event may occur affects asset prices. The event must be difficult, perhaps even impossible, to accurately predict using economic history.' In practice,

https://doi.org/10.1017/S0968565015000025 Published online by Cambridge University Press

this problem creates sample biases in econometric analyses of crises. Indeed, a peso problem occurs when positive probabilities were initially assigned to ultimately missing events. The actual observations used to calibrate or estimate models do not properly account for the real-time evidence, and predictions prove inaccurate.

Second, biases may result from deliberate choices made by researchers. For instance, many economists simply discard wartime data. They implicitly assume that such data are not generated by the same economic mechanisms as 'normal' (i.e. peacetime) data. Be that as it may, focusing on 'normal data' raises the issue of defining 'normality'. The assumption that wartime data is 'abnormal' from a historical standpoint may seem ludicrous. As pointed out by Lerner (1954, p. 506), 'economists have left whole areas of valuable statistical information unexamined because they have considered wartime figures "abnormal"'. Sixty years after this statement, most authors continue to omit wartime observations when dealing with long-term economic series. This can result in severe sample biases, including survivorship bias.

Third, even if all economic crises were intrinsically similar in nature, the current approaches in economics would be too biased and incomplete to address this issue. Indeed, crisis investigations fail to capture the whole universe from which crises are drawn. The typically backward-looking approach economists tend to adopt through econometric investigation lacks imagination about the emergence of new crises that may be the first of their kind. Such new crises can result from new technologies or new ideologies. Within an ever-evolving human society, the shapes of economic crises inevitably change.

Lack of imagination, exclusion of 'abnormal data', and the peso problem may lead to inappropriate understanding of the nature of economic crises.[12] Based on real-life situations, the remainder of this section shows how these three issues reinforce the reference-class problem.

The term 'peso problem' was coined in 1976 following Milton Friedman's comments on the Mexican peso market (Sill 2000). At the time, traders were speculating on a large devaluation of the peso. Accordingly, the currency sold at a discount on the forward market with respect to the spot market. Eventually, however, the peso did not devalue. Hence, spot and forward prices give the wrong impression that the market was inefficient during the speculative period. This biased conclusion is mainly driven by the impossibility of observing traders' real-time expectations.

More generally, a peso problem designates a situation in which traders pay attention to the possible occurrence of a dramatic event that ultimately fails to occur. In such situations, market prices are driven by pessimistic, but most likely rational, expectations. However, ex post there is no observable crisis associated with the price decline, which therefore seems irrational. Ex ante, however, being pessimistic might have made perfect sense. What would have happened if the 'Y2K bug' had materialized or if the Cuban missile crisis had turned into a world war?

[12] See also Witt (2009) on theorizing about novelty.

Empirical findings have been criticized for failing to identify peso problems. Krasker (1980) uses this argument to refute the common belief that the mark/pound exchange rate was irrationally priced during the German hyperinflation of 1922–3. Burnside *et al.* (2011) attribute the claimed profitability of carry trade speculation to a peso problem.[13] Notwithstanding, the economic literature on crises tends to overlook the peso-problem issue. This may be due to the difficulty in determining not only (unrealized) dramatic events but also the moment when traders started to factor them into their expectations.

Economic history shows that identifying peso-problem situations is extremely hard, especially because analysts know the eventual outcome and suffer from an ex post bias. For instance, several papers attempt to understand market anticipations regarding conflict outcomes.[14] In some cases, sharp market movements are not associated with events that historians view as dramatic. According to Willard *et al.* (1996) the most important break in the prices of greenbacks during the US civil war corresponds to the retreat of Confederate General Jubal Early from the suburbs of Washington, an episode largely downplayed by modern historians. Samples of crises built from ex post historical appreciation thus most likely omit crises perceived as such in real time. This is due to interpretations made with knowledge of subsequent events. As pointed out by Frey and Waldenström (2007, p. 3), 'this knowledge may bias the evaluation of the events, and may lead to "facts" being overlooked or overemphasized, as the case may be'.

In sum, peso problems may harm an empirical analysis involving unobservable expectations.[15] They lead researchers to conclusions plagued by missing variables associated with unrealized dramatic events. Circumventing peso problems requires gathering relevant information on real-time circumstances and expectation formation. This is mostly unfeasible for large cross-section studies. In particular, regarding the economic analysis of crises, peso problems raise challenging issues. Should researchers ignore crises that had been feared but never occurred? Otherwise, how should they classify these unborn crises? Since ex post biases are inevitable, peso problems should be addressed from a theoretical perspective.

Economists often discard wartime data, considering it 'abnormal'. But history shows that the 'abnormal' wartime environment can in some instances be the norm. During the period of Louis XIV's reign when he ruled alone (1661–1715), France was at war for more than 60 per cent of the time. For the Napoleonic

[13] Carry trading is a strategy in which the investor borrows in a low-interest-rate currency and lends in a high-interest-rate one. In efficient markets this strategy produces no abnormal returns. Empirically, though, carry trading seems to produce large payoffs that are uncorrelated to traditional risk factors. Nevertheless, Brière and Drut (2009) point out that crises enhance the profitability of carry trades.

[14] Willard *et al.* (1996), Brown and Burdekin (2000, 2002), Weidenmier (2002), Frey and Kucher (2000), Oosterlinck (2003), Frey and Waldenström (2004) and Zussman *et al.* (2008).

[15] Admittedly, this pitfall affects not only economics but all sciences involving human reactions and expectations. But that should not make economists feel better!

https://doi.org/10.1017/S0968565015000025 Published online by Cambridge University Press

period (1800–15), the figure is close to 80 per cent. Figures are even higher in the case of Philip II of Spain: during his reign (1556–98), the empire was at peace for just one year (Drelichman and Voth 2011). Furthermore, although wars usually take place in a well-defined time frame, the beginning and end dates of crises are often blurred and economists fail to reach a consensus on them (Lo 2012). Studies should thus be comprehensive enough to encompass all periods. Omitting 'abnormal' data also prevents an analysis of economic crises in their natural context.

Additionally, disregarding wartime data can trigger a survivorship bias. We illustrate this observation using the so-called 'equity premium puzzle'. In their seminal paper, Mehra and Prescott (1985) argue that the size of the equity premium observed from US equities cannot be explained by standard general equilibrium models unless agents exhibit unrealistically high levels of risk aversion. This puzzle subsequently became a key issue in finance and macroeconomics (Kocherlakota 1996). A convincing solution to this puzzle is proposed by Jorion and Goetzmann (1999), drawing on the survivorship bias associated with sample selection. The authors point out that most empirical studies concentrate on uninterrupted markets only, and equity returns are therefore much higher than if all markets were taken into account. To prove their point, Jorion and Goetzmann (1999) build a database comprising both interrupted and uninterrupted markets. They find that the equity premium is indeed significantly lower in interrupted markets. The deliberate omission of 'abnormal' data may thus be sufficient to create a long-lasting 'puzzle'.

The 1918 repudiation by the Soviets of the bonds issued by the Tsarist regime shows how hard it is to imagine new forms of crises before they occur. This bond crisis most likely represents the first historical example of ideologically motivated debt repudiation.

At the end of the 1890s France allowed Russia to issue a huge number of government bonds on the Paris stock exchange. Political reasons played a key role. As tensions with Germany escalated, the French government wanted to make sure that its Russian ally would remain faithful. The Russian government, knowing the importance of external funding, heavily bribed French politicians (including the finance minister), the head of the stock exchange and the press (Raffalovitch 1931). As a result, by 1914 Russian bonds were trading on the Paris Bourse for an amount estimated to represent a staggering 4.5 per cent of French wealth (Ukhov 2003). When the Soviets managed to take power, they refused to recognize Tsarist debt and decided, in February 1918, to repudiate all former Russian debts. Despite this extreme statement, bondholders kept hoping to be reimbursed. More than two years after the repudiation, some bonds were still traded above 50 per cent of par (Oosterlinck and Landon-Lane 2006).

French investors who had bought Russian bonds were certainly not irrational. When the repudiation occurred they had legitimate reasons to hope they would still be reimbursed. At the time, there was no historical precedent in which a country had, solely for political reasons, repudiated a huge amount of debt without reaching a settlement sooner or later. Even in the dramatic hours following the French Revolution, the legality of the royal debt was not questioned. The revolutionaries defaulted on two-thirds of

the debt but bondholders nonetheless recovered part of their investment (Aftalion 1996). Scholars working on crises in 1917 could not possibly have included the probability of an as-yet unknown form of crisis in their work: the Soviet Revolution. The literature on sovereign debts considers three main reasons why countries would repay their debt: fear of loss of reputation (and thus loss of access to the capital market), fear of military intervention and fear of trade retaliation. The Russian case is exceptional inasmuch as the Soviets did not fear any of these threats and thus the repudiation completely fell out of the range of considered outcomes.

The problems presented in the Russian case are in fact quite common. When the first commodity bubble, centred on tulips, imploded in the seventeenth century, observers were unable to understand that such a crisis could have occurred at all. They blamed wild speculation and irrationality. In the absence of a precedent in which commodity prices would have followed a bubble-shaped pattern, they were at a loss to include such a crisis in their expectations. The same may be said for the South Sea and Mississippi bubbles. In both cases, schemes to exchange government bonds for company shares ended up in a new form of crisis. Even the 1929 stock market crash may be viewed as a totally new form of crisis. World-famous economists such as Irving Fisher completely failed to foresee the crisis, which would eventually lead to a dramatic drop in the prices of all the shares traded on the exchange.[16]

The recent sovereign debt crisis in the Eurozone provides another enlightening example. The risks related to the creation of a single-currency zone with no lender of last resort were hard to fathom. In fact, some scholars believed that the advent of the Eurozone had allowed countries to overcome their 'original sin' problem, i.e. their incapacity to issue debt abroad in their own currency. Eichengreen *et al.* (2005) stressed the positive effects of the euro on sovereign debt issuance. Their measure of original sin for Euro-land countries experienced a steep fall from (0.53 to 0.09) following the introduction of the euro. In contrast, when the crisis occurred, struggling countries proved unable to print more euros to reduce the debt burden.

Economics has been called into question by the 2007–8 financial crisis, which most researchers failed to predict. In our view, this failure is due more to a lack of imagination than to inadequate models. If anything, the blame should be placed on restrictive interpretations of model outcomes, not on the models themselves. In addition, failure to predict crises accurately is not specific to economics; it is also frequent in the natural sciences (Franz 2011).

We argue that the only way to overcome the three issues exemplified in this section is to start from theoretical models comprehensive enough to address the reference-class problem. Such models should allow researchers to use all available data without distinction. While sample biases and peso problems are likely to affect any empirical study, they can be largely avoided when empirical analysis follows from a proper theoretical framework.

[16] Writing in the *New York Times* on 22 Oct. 1929, a few days before Black Thursday, Irving Fisher said 'Stock prices are low.'

https://doi.org/10.1017/S0968565015000025 Published online by Cambridge University Press

V

In the natural sciences, explanatory power and realism are deeply related. The success of theories in physics relies on empirical confirmation, mainly through experimentation. In that context, the explanatory power is bound to the realism of the model: what the model explains *is* the reality under study. Since the models are empirically verifiable or falsifiable, their explanatory capacity merges with a descriptive and predictive capacity. The models concurrently explain and describe the world as it is, and predict how it will be. Inescapably, models evolve with time, but this is the best researchers can do. As Batisty and Domotor (2008, p.169) put it: 'True, a model that works may still prove to be inadequate in some respects in the future. But accepting the mere possibility of such a future discovery does not in any way undermine *our rationality* as scientists in our commitment to the best-working model we currently have.'

For obvious reasons, economists are unable to run crisis experiments. Instead, they need to rely on theoretical models and historical data. When experimentation is unfeasible, the simultaneity of explanation and prediction is possible only if the phenomena under study are stable and repetitive, like the movements of planets, which is not the case for economic crises. Like epidemiologists who have announced the possibility of global pandemics, which ultimately did not materialize (mad cow disease, avian flu), economists find themselves in an uncomfortable situation when dealing with crises (Franz 2011). Weighing the risk of panic against the accusation of being incapable of predicting future crises is a challenging trade-off.

For natural scientists, realism is the most important issue. Unification-type explanations are more suited to the social sciences than are the deductive-nomological explanations of the natural sciences (De Scheemaekere 2009). Explanations indeed increase our understanding of the world by reducing the number of independent phenomena that we have to accept as ultimate. Hence, an economic theory has a significant explanatory capacity insofar as it describes the economic world with fewer independent phenomena – and thus makes this world, other things equal, more comprehensible (Friedman 1974). This is especially relevant when data are unreliable.

It might be of great interest for a theory to account for once-expected large changes that did not ultimately occur (peso problem situations). Given the radical uncertainty underlying financial crises, the balance between the urge to predict (realism) and the need to understand (explanatory value) should tip in favour of the latter, because what matters ultimately are the human decisions taken for reducing the occurrence of crises, not the (impossible) accurate prediction of the next crisis. In economics, the link between a theoretical model and its empirical verification rests on the soundness of human judgement, not the precision of a mechanical measurement.

To this day there is no consensus regarding the definition and classification of crises. Economic studies are deeply rooted in the frequentist approach. Typically, they use an ex post threshold-based definition of crises. This implicitly assumes that all crises are simply tail observations driven by the same underlying probability distribution as the

'normal situation'. This contradicts the intuition that crises exhibit features that make the underlying model behave differently than during 'normal' times. This approach is also plagued by the reference-class problem formulated by Reichenbach ([1935] 1978). Attributing a given crisis to one or another class can lead to dramatically different conclusions, including for policy recommendation.

Focusing on past crises and disregarding crises that were eventually averted is akin to building a medical science by ex post examining incurable patients only, not those who either did not get sick or were successfully cured in due course. Moreover, economists lack imaginative models to apprehend new forms of crises. Rather, they tend to restrict their models to either 'normal' or 'abnormal' past periods. This distinction ignores the fact that reality is a succession of both normal and crisis periods. In particular, crises start in normal periods. Whereas in the natural sciences, models can predict the existence of particles before they are empirically discovered, such as the Brout–Englert–Higgs boson, or diseases before their outbreak, like disease mutations, existing economic models fail to consider potential innovations in crises. Potential crises come from the theory itself, not from ex post – and typically ad hoc – considerations. Models that are supposed to hold only under 'normal conditions', leaving the possibilities for crises aside, are too narrowly conceived.

Submitted: 23 June 2014
Revised version submitted: 21 February 2015
Accepted: 28 February 2015

References

ACCOMINOTTI, O., FLANDREAU, M. and REZZIK, R. (2011). The spread of empire: Clio and the measurement of colonial borrowing costs. *Economic History Review*, **64**, pp. 385–407.

AFTALION, F. (1996). *L'économie de la Révolution française*. Paris: PUF, Collection Quadrige.

ALLEN, W. A. and MOESSNER, R. (2012). The international propagation of the financial crisis of 2008 and a comparison with 1931. *Financial History Review*, **19**, pp. 123–47.

BARKBU, B., EICHENGREEN, B. and MODY, A. (2012). Financial crises and the multilateral response: what the historical record shows *Journal of International Economics*, **88**, pp. 422–35.

BATISTY, V. and DOMOTOR, Z. (2008). Chaos, complexity, and indeterminism. In J. V. Ciprut (ed.), *Indeterminacy: The Mapped, the Navigable, and the Uncharted*, Cambridge, MA: MIT Press.

BAUMAN, P. (2005). Three doors, two players and single-case probabilities. *American Philosophical Quarterly*, **42**, pp. 71–9.

BAUMAN, P. (2008). Single-case probabilities and the case of Monty Hall: Levy's view. *Synthese*, **162**, pp. 265–73.

BERG, A. and PATTILLO, C. (1999). Are currency crises predictable? A test. *IMF Staff Papers*, **46**, pp. 107–38.

BERNAL, O., OOSTERLINCK, K. and SZAFARZ, A. (2010). Observing bailout expectations during a total eclipse of the sun. *Journal of International Money and Finance*, **29**, pp. 1193–205.

BRIERE, M., CHAPELLE, A. and SZAFARZ, A. (2012). No contagion, only globalization and flight to quality. *Journal of International Money and Finance*, **31**, pp. 1729–44.

BRIERE, M. and DRUT, B. (2009). The revenge of purchasing power parity on carry trades during crises. *CEB Working Paper* no. 09/013.

BROWN, W. O. and BURDEKIN, R. C. K. (2000). Turning points in the US civil war: a British perspective. *Journal of Economic History*, **60**, pp. 216–31.

BROWN, W. O. and BURDEKIN, R. C. K. (2002). German debt traded in London during World War II: a British perspective on Hitler. *Economica*, **69**, pp. 655–69.
BURNSIDE, C., EICHENBAUM, M., KLESHCHELSKI, I. and REBELO, S. (2011). Do peso problems explain the returns to the carry trade? *Review of Financial Studies*, **24**, pp. 853–91.
CANDELON, B., DUMITRESCU, E.-I. and HURLIN, C. (2012). How to evaluate an early-warning system: toward a unified statistical framework for assessing financial crises forecasting methods. *IMF Economic Review*, **60**, pp. 75–113.
CHANG, R. and VELASCO, A. (2001). A model of financial crises in emerging markets. *Quarterly Journal of Economics*, **116**, pp. 489–517.
DAVIES, W. and MCGOEY, L. (2012). Rationalities of ignorance: on financial crisis and the ambivalence of neo-liberal epistemology. *Economy and Society*, **41**, pp. 64–83.
DE SCHEEMAEKERE, X. (2009). The epistemology of modern finance. *Journal of Philosophical Economics*, **2**, pp. 99–120.
DE SCHEEMAEKERE, X. and SZAFARZ, A. (2009). The special status of mathematical probability: a historical sketch. *Epistemologia*, **32**, pp. 91–110.
DRELICHMAN, M. and VOTH, H.-J. (2011). Serial defaults, serial profits: returns to sovereign lending in Habsburg Spain, 1566–1600. *Explorations in Economic History*, **48**, pp. 1–19.
EAGLE, A. (2004). Twenty-one arguments against propensity analyses of probability. *Erkenntnis*, **60**, pp. 371–416.
EATON, J. and FERNANDEZ, R. (1995). Sovereign debt. *NBER Working Paper* 5131.
EICHENGREEN, B., HAUSSMANN, R. and PANIZZA, U. (2005). The pain of original sin. In B. Eichengreen and R. Hausmann (eds.), *Other People's Money: Debt Denomination and Financial Instability in Emerging Market Economies*. Chicago: University of Chicago Press.
EICHENGREEN, B. and PORTES, R. (1987). The anatomy of financial crises. *NBER Working Paper* 2126.
FERGUSON, N. and SCHULARICK, M. (2006). The empire effect: the determinants of country risk in the first age of globalization, 1880–1913. *Journal of Economic History*, **66**, pp. 283–312.
FLANDREAU, M. (2012). The vanishing banker. *Financial History Review*, **19**, pp. 1–19.
FRANKEL, J. and SARAVELOS, G. (2012). Can leading indicators assess country vulnerability? Evidence from the 2008–09 global financial crisis. *Journal of International Economics*, **87**, pp. 216–31.
FRANZ, W. (2011). Fear mongering. *ZEW News*, 12.
FREY, B. and KUCHER, M. (2000). History as reflected in capital markets: the case of World War II. *Journal of Economic History*, **60**, pp. 468–96.
FREY, B. and WALDENSTRÖM, D. (2004). Markets work in war: World War II reflected in Zurich and Stockholm bond markets. *Financial History Review*, **11**, pp. 51–67.
FREY, B. and WALDENSTRÖM, D. (2007). Using financial markets to analyze history: the case of the Second World War. *Historical Social Research*, **32**, pp. 330–50.
FRIEDMAN, M. (1974). Explanation and scientific understanding. *Journal of Philosophy*, **71**, pp. 5–19.
HÁJEK, A. (1997). 'Mises redux' – redux: fifteen arguments against finite frequentism. *Erkenntnis*, **45**, pp. 209–27.
HÁJEK, A. (2007). The reference class problem is your problem too. *Synthese*, **156**, pp. 563–85.
JORION, P. and GOETZMANN, W. (1999). Global stock markets in the twentieth century. *Journal of Finance*, **54**, pp. 953–80.
KAMINSKY, G., LIZONDO, S., and REINHART, C. M. (1998). Leading indicators of currency crises. *IMF Staff Papers*, **45**, pp. 1–48.
KNIGHT, F. H. (1921). *Risk, Uncertainty and Profit*. Boston and New York: Houghton Mifflin Company.
KOCHERLAKOTA, N. R. (1996). The equity premium: it's still a puzzle. *Journal of Economic Literature*, **34**, pp. 42–71.
KRASKER, W. S. (1980). The 'peso problem' in testing the efficiency of forward exchange markets. *Journal of Monetary Economics*, **6**, pp. 269–76.
LERNER, E. (1954). The monetary and fiscal programs of the Confederate government, 1861–1865. *Journal of Political Economy*, **62**, pp. 506–22.
LEVY, K. (2007). Bauman on the Monty Hall problem and single-case probabilities. *Synthese*, **158**, pp. 139–51.

LO, A. W. (2012). Reading about the financial crisis: a twenty-one-book review. *Journal of Economic Literature*, **50**, pp. 151–78.
MEHRA, R. and PRESCOTT, E. C. (1985). The equity premium: a puzzle. *Journal of Monetary Economics*, **15**, pp. 145–61.
MILLSTEIN, R. L. (2003). Interpretations of probability in evolutionary theory. *Philosophy of Science*, **70**, pp. 1317–28.
MILNE, P. (1986). Can there be a realist single-case interpretation of probability? *Erkenntnis*, **25**, pp. 129–32.
MORGAN, M. S. (1990). *The History of Econometrics Ideas*. Cambridge: Cambridge University Press.
OBSTFELD, M. and TAYLOR, A. M. (2003). Sovereign risk, credibility and the gold standard: 1870–1913 versus 1925–1931. *Economic Journal*, **113**, pp. 241–75.
OOSTERLINCK, K. (2003). The bond market and the legitimacy of Vichy France. *Explorations in Economic History*, **40**, pp. 327–45.
OOSTERLINCK, K. and LANDON-LANE, J. (2006). Hope springs eternal: French bondholders and the Soviet Repudiation (1915–1919). *Review of Finance*, **10**, pp. 507–35.
PIERCE, A. (2008). The Queen asks why no one saw the credit crunch coming. *The Telegraph* online. www.telegraph.co.uk/news/uknews/theroyalfamily/3386353/The-Queen-asks-why-no-one-saw-the-credit-crunch-coming.html
PORTES, R. (1999). An analysis of financial crisis: Lessons for the international financial system. In W. C. Hunter, G. G. Kaufman and T. H. Krueger (eds.), *The Asian Financial Crisis: Origins, Implications, and Solutions*. New York: Springer Science and Business Media New York.
RADELET, S. and SACHS, J. (1998). The East Asian financial crisis: diagnosis, remedies, prospects. *Brookings Papers on Economic Activity*, **1**, pp. 1–74.
RAFFALOVITCH, A. (1931). *L'abominable vénalité de la presse, d'après les documents des archives russes (1897–1917)*. Paris: Librairie du Travail.
REICHENBACH, H. ([1935] 1978). *The Theory of Probability*. In M. Reichenbach and R. S. Cohen (eds.), *Hans Reichenbach, Selected Writings, 1909–1953*, vol. 2. Dordrecht: Reidel.
REINHART, C. and ROGOFF, K. (2004). The modern history of exchange rate arrangements: a reinterpretation. *Quarterly Journal of Economics*, **119**, pp. 1–48.
REINHART, C. and ROGOFF, K. (2011). From financial crash to debt crisis, *American Economic Review*, **101**, pp. 1676–706.
SCARDOVI, I. (1988). Ambiguous uses of probability. In E. Agazzi (ed.), *Probability in the Sciences*. Dordrecht: Kluwer.
SCHULARICK, M. and TAYLOR, A. M. (2012). Credit booms gone bust: monetary policy, leverage cycles, and financial crises, 1870–2008. *American Economic Review*, **102**, pp. 1029–61.
SETTLE, T. (1977). Popper versus Peirce on the probability of single cases. *British Journal for the Philosophy of Science*, **28**, pp. 177–80.
SILL, K. (2000). Understanding asset values: stock prices, exchange rates, and the 'peso problem'. *Business Review*, September/October, pp. 3–13.
SVETLOVA, E. and FIEDLER, M. (2011). Understanding crisis: on the meaning of uncertainty and probability. In O. Dejuãn, E. Febrero and M. C. Marcuzzo (eds.), *The First Great Recession of the 21st Century*. Cheltenham, UK: Edward Elgar.
SVETLOVA, E. and VAN HELST, H. (2015). Decision-theoretic approaches to non-knowledge in economics. In M. Gross and L. McGoey (eds.), *Routledge International Handbook of Ignorance Studies*. London: Routledge.
SZAFARZ, A. (1984). Richard von Mises: l'échec d'une axiomatique. *Dialectica*, **38**, pp. 311–17.
SZAFARZ, A. (2009). How did financial-crisis-based criticisms of market efficiency get it so wrong? *Working Paper CEB 9–48*.
SZAFARZ, A. (2012). Financial crises in efficient markets: how fundamentalists fuel volatility. *Journal of Banking & Finance*, **36**, pp. 105–11.
TALEB, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.
TEMIN, P. (2008). The German crisis of 1931: evidence and tradition. *Cliometrica*, **2**, pp. 5–17.
UKHOV, A. D. (2003). Financial innovation and Russian government debt before 1918. *Yale Working Paper* no. 3–20.

VERONESI, P. (2004). The peso problem hypothesis and stock market returns. *Journal of Economic Dynamics and Control*, **28**, pp. 707–25.
WEIDENMIER, M. (2002). Turning points in the US Civil War: views from the grayback market. *Southern Economic Journal*, **68**, pp. 875–90.
WILLARD, K., GUINNANE, T. and ROSEN, H. (1996). Turning points in the Civil War: views from the greenback market. *American Economic Review*, **86**, pp. 1001–18.
WITT, U. (2009). Novelty and the bounds of unknowledge in economics. *Journal of Economic Methodology*, **16**, pp. 361–75.
ZECKHAUSER, R. J. and VISCUSI, W. K. (1990). Risk within reason. *Science*, **248**, pp. 559–64.
ZUSSMAN, A., ZUSSMAN, N. and NIELSEN, M. O. (2008). Asset market perspectives on the Israeli–Palestinian conflict. *Economica*, **75**, pp. 84–115.