purchases from the largest seller remain high until 2005, where the largest contract is twice as large as it was in 1999. After 2005, the quantity sold on the largest contract begins to decline for deregulated utilities, coincident with the rise in contract prices shown in panel (a). These figures are consistent with contracts at low prices expiring around 2005 and being replaced by more expensive ones.

6.2 A Case Study of Delayed Effective Deregulation: Illinois

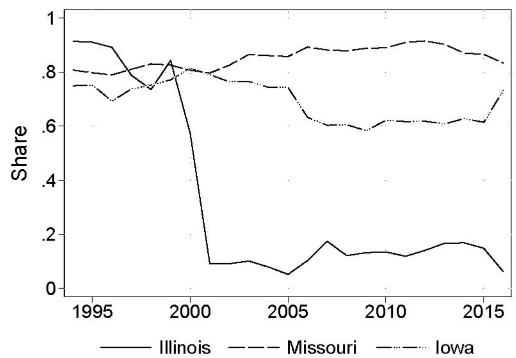
To help illustrate how the timing of deregulation was delayed by state-specific measures, we present Illinois as a case study. In the 1990s, Illinois' electricity rates were among the highest in the United States. Motivated by these high prices, Illinois lawmakers passed the Consumer Choice Act in 1997, which encouraged large investor-owned utilities to divest their generation assets and allowed for independent companies to supply electricity to commercial customers. For residential customers and small businesses, rates were lowered by 15 percent and frozen for 10 years. In 2002, retail choice was extended to residential and small commercial customers, thus allowing for competitive supply in the downstream market.

Within a few years, the investor-owned utilities in Illinois had sold off their complete portfolio of generation assets. This large change to the market is illustrated in panel (a) of Figure 12. The solid black line represents the share of sources that investor-owned utilities obtained from their own generation. The remainder is obtained by purchasing electricity from other producers. The share of electricity sourced from own generation fell from above 80 percent at the time of the restructuring initiatives to 10 percent by 2001.

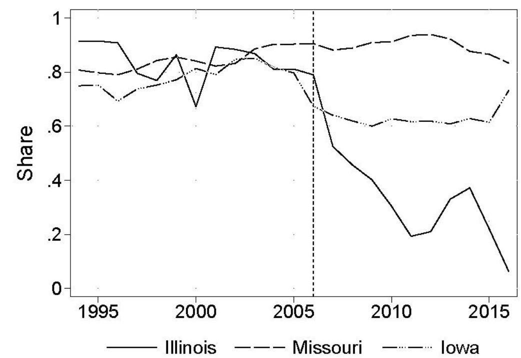
For comparison, we construct two reference groups: (1) investor-owned utilities in Missouri and (2) investor-owned utilities in Iowa. Missouri is a neighboring state and its largest utility, Union Electric, is part of the Ameren group that owns the utilities serving a large portion of Illinois. Iowa is also a neighboring state, and its largest utilities serve part of northwest Illinois. Importantly, neither Missouri nor Iowa passed any deregulation measures in this period. Panel (a) of Figure 12 plots the share of own generation for Missouri utilities in a dashed line and for Iowa utilities in a dash-dot line. While deregulated firms in Illinois divested nearly all of their generation assets, the regulated firms in Missouri and Iowa continued to obtain the vast majority of their electricity from own generation.

Even though deregulated firms legally divested themselves of generation assets quickly, the actual restructuring of the upstream market came about more slowly. Panel (b) of Figure 12 plots the share of electricity obtained from affiliated companies, which combines both own generation and purchases from companies belonging to the same parent company. The share of purchases from affiliated companies did not fall until 2007. In practice, Illinois utilities split into subsidiary companies and signed long-term purchase agreements with each other at the time of divestiture. The last year of these contracts (2006) is indicated by the vertical dashed line. Even at the end of the sample, some fraction of the electricity is still purchased from

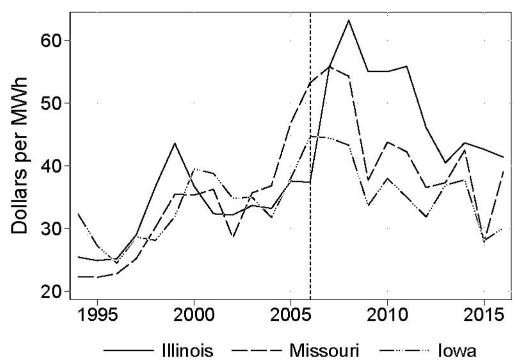
Figure 12: Timing of Deregulation: Illinois



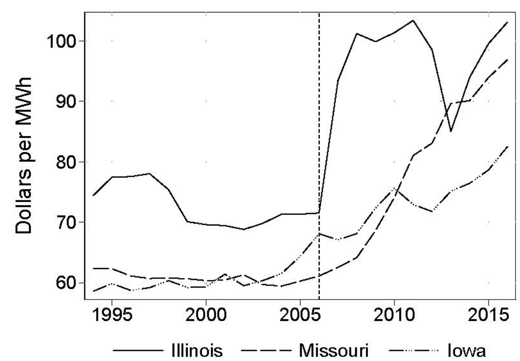
(a) Share of Electricity from Own Generation



(b) Share of Electricity from Affiliated Sources



(c) Upstream (Wholesale) Prices



(d) Downstream (Retail) Prices

Notes: Panel (a) of the figure displays the share of incumbent utilities' total sources provided by own generation for Illinois, which deregulated, and Missouri and Iowa, which did not. Panel (b) plots the share of incumbent utilities' total sources provided by affiliated sources, which include both own generation and purchases from companies belonging to the same parent company. Panels (c) and (d) display average wholesale purchase prices and retail prices, respectively. The year 2006, which is the final year of several long-term contracts between affiliated companies, is indicated by a vertical dashed line.

affiliated companies, raising the possibility that aspects of vertical integration might still be at play in the market.

In the downstream market, consumers were slow to switch from the incumbent utilities due to the price caps that kept utility rates low. The price cap on rates expired in 2007, and many customers switched to independent retailers in that year. Thus, effective deregulation, measured by the impact on market restructuring, did not occur in Illinois until roughly 2007, when most wholesale transactions were between independent parties and retail choice became much more common.

Though deregulation was expected to bring down prices, wholesale electricity prices in Illinois increased sharply in 2007, when contracts expired and deregulation had effectively taken place. This is illustrated in panel (c) of Figure 12. Before 2007, the quantity-weighted purchase price for deregulated utilities in Illinois followed a similar path to prices in Missouri and Iowa. After effective deregulation, wholesale prices in Illinois spiked, and then stayed above prices paid by regulated utilities.

Panel (d) of Figure 12 plots the downstream retail prices. The solid line in the plot shows that prices were steady from 1999 through 2006, which corresponds to the period that the rate freeze was in effect. At the expiration of the rate freeze, retail prices spiked. This increase was sudden and large relative to the price patterns observed in Missouri and Iowa.

Note that, according to our calculations of the change in consumer surplus in Table A2, Illinois is the state that benefited the most from deregulation. On average, consumers in Illinois realized *lower* prices than those charged by comparable utilities. In fact, despite the large initial jump in prices after the rate freeze was removed, Figure 12 shows that retail prices in Iowa and Missouri increased at a faster rate than those in Illinois. Until 2010, the figure suggests that Illinois had significantly higher prices than its neighbors.

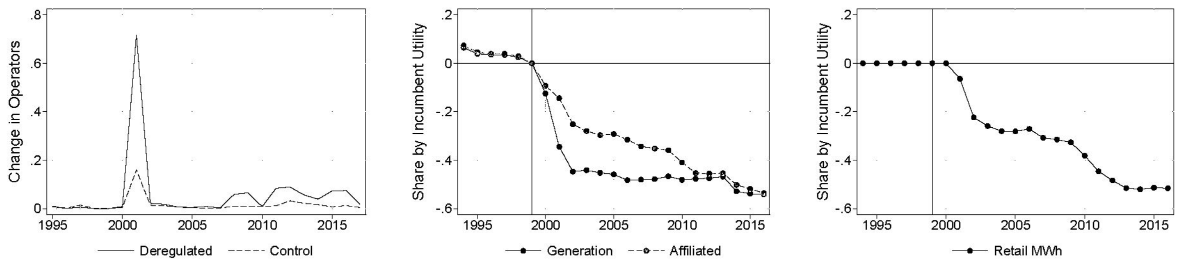
The case study of Illinois illustrates how the effects of deregulation can be delayed for several years, even when legal measures such as vertical separation and competitive markets are introduced quickly. Firms have access to mechanisms (e.g., contracts and umbrella ownership) to maintain a strong degree of vertical integration even when legal entities are vertical separated. Even though Illinois is the most consumer-friendly scenario according to our analysis, wholesale and retail prices increased significantly around the time of effective deregulation.

6.3 Aggregate Delays in Effective Deregulation

Here, we present the estimated delays arising across all deregulated utilities in our sample. First, in panel (a) of Figure 13, we plot the share of generation that reported a new operator from the previous year. Consistent with the narrative of divestiture, approximately 70 percent of generated MWh was under a new operator in 2001. This event is an extreme outlier in the graph, as no more than 10 percent change operators outside of 2000–2002. Next, we consider the difference-in-difference estimates for shares of the incumbent utility. Panel (b) shows our measure of effective deregulation in the upstream market. The solid black line shows the change in the share of aggregate retail consumption that was generated by incumbent utilities. The generation shares fell steeply from 1999 to 2002, with a drop of 44 percentage points. A few additional separations occurred in later years, with the total decline in generation shares reaching 54 percentage points by 2016. We do not observe a decline of 100 percentage points for two reasons. First, deregulated utilities were obtaining only roughly 80 percent of their consumed electricity in 1999 from generation, providing an upper bound for the effect of deregulation. Second, not all utilities in deregulated states were forced to separate generation from retail. For example, in Texas, only IOUs in the ERCOT region were affected. The other IOUs continued to operate as vertically integrated entities.

The grey dashed line shows the affiliated generation share, which captures all generation

Figure 13: Apparent versus Effective Deregulation



(a) Generation Operators (b) Upstream Utility Share (c) Downstream Utility Share

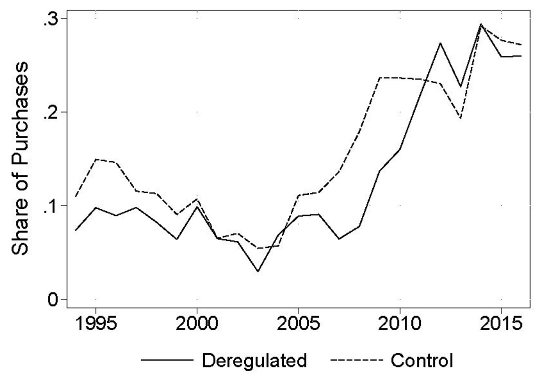
Notes: Figure shows changes in upstream and downstream markets after deregulation. Panel (a) plots the raw share of generation that changed operators from one year to another. Panels (b) and (c) present difference-in-differences matching estimates of changes in the incumbent utility's share of the upstream wholesale market and the downstream retail market. Panel (a) plots a utility's share of quantity demanded provided by its own generation and by all affiliated sources. The gap between the two lines indicates a delay between apparent deregulation and effective deregulation attributable to contracts and umbrella ownership. Panel (b) shows the change in the incumbent utility's share in the downstream retail market.

occurring from utilities and generators owned by the same parent companies. This measure proxies for the long-term contracts signed my several utilities with their generators at the time of separation. The grey dashed line shows that the actual changes to the wholesale market lagged the apparent changes for many years. Though the naive share of competitive generation (i.e., one minus the point estimates in the graph) had increased by over 40 percentage points in 2002, this actual share of competitive generation did not cross this threshold until 2010, after accounting for umbrella ownership across generators and utilities. By 2011, our measures converge, which is consistent with the expiration of the initial contracts and the completion of the transition to a competitive wholesale market.

This narrative lines up with the changes in costs we observe in Figure 5. From 2000 through 2004, while many of these contracts were in effect, generation costs and wholesale costs barely changed. Coincident with the decline in affiliated generation shares starting in 2005, generation costs fell and wholesale markups increased. Taken together, these patterns are consistent with utilities signing long-term contracts at prevailing rates with their separated generation facilities, which delayed the onset of competitive markets for many years. The timing of these cost increases contribute to the larger increases in prices we observe starting in 2006.

A second restriction that delayed the onset of competitive retail markets was the practice of implementing retail rate freezes in deregulated states. These rate freezes kept retail prices low, making the existing utility attractive to consumers and effectively discouraging new entrants. As shown in panel (b) of Figure 5, deregulated utilities saw a decrease in retail markups from 2000 to 2008. These rate freezes could have delayed the transition to competitive retail markets. As shown in panel (b) of Figure 13, competitive retailers obtained roughly 30 percent of the market by 2003. The transition plateaued at this level for several years. Beginning in 2007, the retail

Figure 14: Share of Purchases from ISOs and Power Pools



Notes: Figure displays the shares of purchased electricity obtained from ISO wholesale markets and power pools, for utilities in deregulated and control states. The residual shares are from bilateral contracts with electricity suppliers.

market saw a gradual increase in competitive providers, reaching 52 percent of the market by 2016.

7 Possible Alternative Explanations

In this section, we discuss other events that had an impact on electricity prices and costs that could potentially play a role explaining our findings. Overall, we find that the weight of the evidence points the substantial role of market power in explaining the increase in prices and markups that we observe in deregulated states.

7.1 ISO Markets

During the restructuring process, transmission assets covering areas much larger than a single utility's service area were put into the hands of an independent operator. This served two purposes: First, to grant easier access to independent generators who wanted to sell energy into the market. Second, to allow for trade across larger areas as a potential channel to reduce costs by sourcing energy from low cost plants. Evidence indicates that central dispatch by regional transmission operators has indeed reduced costs (Cicala, 2022).

Our findings suggest that market power in the wholesale market started increasing shortly after ISO organized markets started operating.[41] Nonetheless, there are several reasons why the opening of centrally dispatched electricity markets is unlikely to lead to the observed increase in market power. First, we would expect ISO markets to strengthen competition rather than weaken it, since they connect a larger number of players and have transparent market clearing prices. Second, even if they increase generators' market power, the share of electricity that

[41] For example, the Midwest market (MISO) started operating in 2005 and the New England ISO in 2004.

utilities purchased from ISOs in those early years was fairly low. Figure 14 plots the share of purchased power coming from ISO markets and power pools, which were the predecessors to ISOs. Before 2008, sales via these centralized markets accounted for less than 15 percent of overall wholesale transactions. As of 2016, the vast majority of all purchased power was through traditional bilateral contracts—not ISOs.[42]

Further, ISO markets are not exclusive to restructured states. For instance, only 2 of the 10 states belonging to MISO in 2005 were restructured, and MISO is the second largest ISO after PJM. Figure 14 shows that the share of purchased electricity from ISOs was roughly similar across deregulated and control states. Since our analysis compares utilities in restructured and regulated states, we think that it is unlikely that the observed difference in market power would come from ISO purchases.

Finally, the increase in wholesale markups we observe is not restricted to one market mechanism. Deregulated states see relative increases in both spot market prices (from ISOs and power pools) and bilateral contract prices. Though prices in spot markets tend to track marginal costs more closely, the high-level patterns are similar in the bilateral contract market. Figure A7 in the appendix plots the average purchase prices for deregulated and control utilities from ISOs and power pools compared to bilateral contracts. The plots indicate that wholesale prices in deregulated states realized relative increases of roughly 10 dollars per MWh from both spot markets and bilateral contracts.

7.2 Stranded Costs

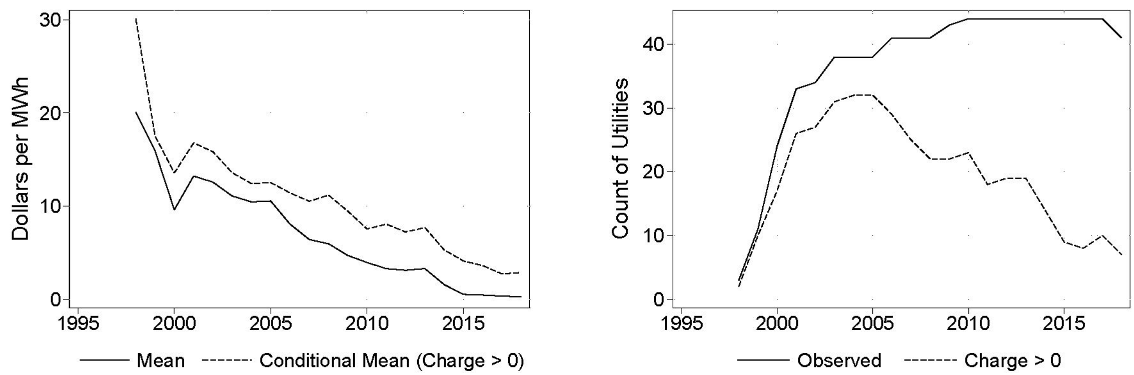
During restructuring, most utilities reached agreements with state regulatory authorities to levy additional charges on their customers related to the move toward deregulation. A common argument by the utilities was that the transition to competitive markets would result in a loss in value of their capital investments, and that they should be compensated for the "stranded" costs of these assets. One question is whether the observed increase in rates reflects these additional charges.

We collected information on transition charges, which covered stranded costs, for 44 large utilities across 16 states that passed deregulation measures.[43] Most of the utilities for which we obtained data levied additional transition charges on their customers; only 6 of them never implemented transition charges. Transition charges were initially very high and decline throughout our sample period. Panel (a) in Figure 15 shows the mean of these additional charges over time. This decline holds even if we condition the mean on utilities with positive stranded costs in each period, thus dropping utilities as their window for stranded cost recovery ends. As shown in panel (b), individual utilities phase out stranded costs starting in 2006. The solid line

[42] If we also account for own generation, the share from ISOs is even smaller. The share from own generation is larger in control states.

[43] The data were obtained from utility ratebooks or the relevant state regulatory commission.

Figure 15: Transition Charges and Stranded Costs



(a) Mean Transition Charges

(b) Count in Sample

Notes: Figure displays the transition charges levied on customers in deregulated utilities to cover stranded costs and other features of restructuring. Panel (a) plots the mean charges (solid line) and the condition mean for positive charges (dashed line). Panel (b) plots the count of utilities with reported transition charges (solid line) and the count of utilities with positive charges (dashed line).

shows the count of utilities for which we have stranded costs measures, and the dashed line shows the count of utilities with positive costs.

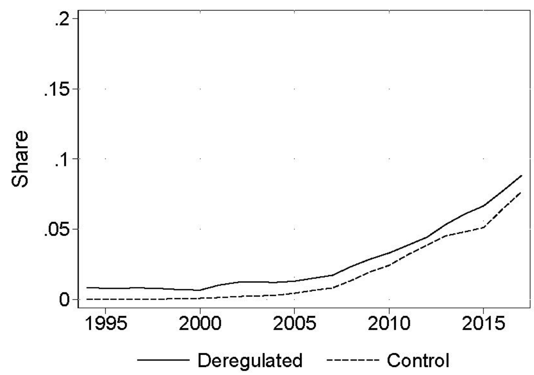
Thus, coinciding with the time we observe effective deregulation and large markup increases, we observe declines in stranded costs and transition charges, with many utilities phasing them out altogether. Though we do not have a complete panel of all stranded costs, we find it very unlikely that they account for the observed increase in prices in deregulated states. The trends in stranded costs move in the opposite direction from the price changes we observe; if anything, these costs may mask some of market power effects of deregulation.

7.3 Renewable Portfolio Standards

Renewable portfolio standards (RPS) require utilities to procure a minimum share of the electricity they sell from renewable sources. RPS have the potential to increase prices (Greenstone and Nath, 2020) and might have contributed to increase utilities' costs, since 25 states had passed regulation with this kind of requirement by 2007.

We think RPS are an unlikely explanation for our results for at least two reasons. First, although RPS were more common among restructured states, those that remained regulated adopted them as well. For example, we find that markups and prices started to diverge around 2006. In 2007, 14 restructured states and 7 regulated states had adopted RPS (Greenstone and Nath, 2020). Second, despite RPS adoption being more likely in deregulated states, the gradual increase in share of generation coming from renewable sources has been similar across the two groups. A possible explanation for this is that at the point of adoption, the requirements put

Figure 16: Share of Generation from New Renewable Resources



Notes: Figure displays share of generated electricity from renewable resources in deregulated and control states. The plot reflects wind, solar, and geothermal sources. Hydropower is excluded because RPS requirements have had little impact on hydropower sources.

in place by RPS were not stringent. To illustrate this, Figure 16 shows the share of generation coming from renewable resources—wind, solar, and geothermal—in deregulated and control states.[44] The figure shows that the shares are nearly identical across the two groups, and they increase at the same gradual rate starting in 2008.

7.4 Other Cost Shocks

Since the restructuring process started, the electric industry has received several cost shocks from two main sources: fuel prices and environmental regulation. How these shocks affected a utility's cost structure depends on the utility's initial fuel mix since, for instance, more stringent environmental regulation will have a stronger effect on costs for utilities that rely more heavily on coal to produce electricity. A potential concern would then be that this initial difference in fuel mix determined how firms were affected by cost shocks, and not the restructuring process.[45]

Our matching approach allows us to deal with this concern, since each utility in a restructured state is compared to utilities in regulated states with a similar fuel mix in 1994. Though our empirical approach compares utilities that in principle would be similarly affected by these cost shocks, it remains vulnerable to variation coming from changes in the fuel mix that took place after 1994. We do not necessarily want to control for entry and exit decisions that took place after the deregulation process had started, as these decisions may have been caused by the deregulation process. If, for instance, deregulated markets attracted more entry by cleaner

[44] Hydropower is excluded because hydropower plants were not the target of RPS requirements. From 2001 through 2016, the share of hydropower generation has remained roughly flat across deregulated and control states.

[45] For example, the Energy Policy Act of 2005 introduced several subsidies and environmental requirements at the federal level, which had varying effects on different types of generators.

plants, or by gas plants that could take advantage of the cheaper gas, this is something that we we can include in our estimates of cost efficiencies. In our data, we observe similar trends in aggregate generation by fuel types across the two groups.[46]

A related concern is that plants may choose emissions compliance strategies that differentially affect their cost structures. Fowlie (2010) compares compliance strategies between restructured and regulated coal plants in response to an emissions trading program introduced in 2006 to regulate NO_X, an ozone precursor. The program affected plants in 19 states, of which 12 were restructured. Because rate-of-return regulation creates stronger incentives for capital investment, regulated plants chose more capital intensive compliance options than plants in restructured states. This implies that environmental regulation could potentially have increased fixed cost for regulated plants and variable costs for restructured plants. If compliance raises variable costs that we do not measure, we could potentially overstate the changes in markups in restructured states. Despite this, compliance costs would not likely explain the large magnitudes that we observe. Engineering estimates of operating compliance costs taken from Fowlie (2010) indicate that the maximum difference between common compliance technologies is around 7.5 dollars per MWh, which is much less than the markup increases that we find (see Figure 4). Moreover, these costs are not much more than the decrease in fuel cost in restructured utilities over that period (see Section 4.3). Thus, such regulations are not likely to generate large increases in variable costs in restructured states.

8 Conclusion

We present a detailed analysis of the evolution of electricity prices and costs from 1994 until 2016. Our analysis spans the implementation of state-specific deregulation measures that began in the late 1990s, which included the introduction of market-based prices. Compared to utilities in states that stayed regulated, deregulated utilities realized higher prices but lower average and marginal costs. Overall, markups increased substantially. Our findings are consistent with the exercise of market power in deregulated markets, particularly at the wholesale level. Generation facilities were able to charge prices at substantial markups above costs, and the vertical separation of generation and retail allowed for additional price increases due to double marginalization.

For our analysis, we construct a unique firm-level dataset that includes firm-to-firm transactions and umbrella ownership that links subsidiaries to the same parent/holding company. We find that changes in prices and markups increased over time because long-term contracts and umbrella ownership delayed the intended changes in vertical market structure. Thus, our research highlights the importance of accounting for intermediate degrees of vertical integration

[46] The only meaningful difference in our data is that control states became relatively less reliant on coal and more reliant on natural gas during our sample period.

to understand the consequences of deregulation and related policies.

Our findings do not necessarily imply that electricity markets should remain regulated, but rather emphasizes the importance of careful oversight of deregulated markets and the consideration of market power in market design. Further research is needed on how to organize markets such that consumers can benefit from lower prices, as well as understanding the longer-run effects of deregulation that arise from changes in investment and environmental compliance efforts.

References

ABADIE, A. AND G. W. IMBENS (2006): “Large Sample Properties of Matching Estimators for Average Treatment Effects,” *Econometrica*, 74, 235–267.

ABADIE, A. AND G. W. IMBENS (2008): “On the Failure of the Bootstrap for Matching Estimators,” *Econometrica*, 76, 1537–1557.

ABADIE, A. AND G. W. IMBENS (2011): “Bias-Corrected Matching Estimators for Average Treatment Effects,” *Journal of Business & Economic Statistics*, 29, 1–11.

BERTRAM, G. AND D. TWADDLE (2005): “Price-Cost Margins and Profit Rates in New Zealand Electricity Distribution Networks Since 1994: The Cost of Light Handed Regulation,” *Journal of Regulatory Economics*, 27, 281–308.

BORENSTEIN, S. (1989): “Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry,” *The RAND Journal of Economics*, 344–365.

——— (2002): “The Trouble with Electricity Markets: Understanding California’s Restructuring Disaster,” *Journal of Economic Perspectives*, 16, 191–211.

BORENSTEIN, S. AND J. BUSHNELL (2000): “Electricity Restructuring: Reregulation or Reregulation?” *Regulation*, 23, 46.

——— (2015): “The US Electricity Industry after 20 Years of Restructuring,” *Annual Review of Economics*, 7, 437–463.

BORENSTEIN, S., J. B. BUSHNELL, AND F. A. WOLAK (2002): “Measuring Market Inefficiencies in California’s Restructured Wholesale Electricity Market,” *American Economic Review*, 92, 1376–1405.

BORENSTEIN, S. AND N. L. ROSE (1994): “Competition and Price Dispersion in the U.S. Airline Industry,” *Journal of Political Economy*, 102, 653–683.

——— (2014): “How Airline Markets Work? Or Do They? Regulatory Reform in the Airline Industry,” in *Economic Regulation and Its Reform: What Have We Learned?*, University of Chicago Press, 63–135.

BURKE, P. J. AND A. ABAYASEKARA (2018): “The Price Elasticity of Electricity Demand in the United States: A Three-Dimensional Analysis,” *The Energy Journal*, 39.

BUSHNELL, J., E. MANSUR, AND K. NOVAN (2017): “Review of the Economics Literature on U.S. Electricity Restructuring,” Working paper.

BUSHNELL, J., E. MANSUR, AND C. SARAVIA (2008): “Vertical Arrangements, Market Structure and Competition: An Analysis of Restructured U.S. Electricity Markets,” *American Economic Review*, 98, 237–66.

CICALA, S. (2015): “When Does Regulation Distort Costs? Lessons from Fuel Procurement in U.S. Electricity Generation,” *American Economic Review*, 105, 411–44.

——— (2022): "Imperfect Markets versus Imperfect Regulation in US Electricity Generation," *American Economic Review*, 112, 409–41.

COASE, R. H. (1960): "The Problem of Social Cost," *Journal of Law and Economics*, 3, 1–44.

DAVIS, L. W. AND C. WOLFRAM (2012): "Deregulation, Consolidation, and Efficiency: Evidence from U.S. Nuclear Power," *American Economic Journal: Applied Economics*, 4, 194–225.

DERYUGINA, T., A. MACKAY, AND J. REIF (2019): "The Long-Run Dynamics of Electricity Demand: Evidence from Municipal Aggregation," *American Economic Journal: Applied Economics*, forthcoming.

EDISON ELECTRIC INSTITUTE (2019): "Profiles and Rankings of Investor-Owned Electric Companies," Technical report.

FABRIZIO, K. R., N. L. ROSE, AND C. D. WOLFRAM (2007): "Do Markets Reduce Costs? Assessing the Impact of Regulatory Restructuring on U.S. Electric Generation Efficiency," *American Economic Review*, 97, 1250–1277.

FAN, S. AND R. J. HYNDMAN (2011): "The Price Elasticity of Electricity Demand in South Australia," *Energy Policy*, 39, 3709–3719.

FOWLIE, M. (2010): "Emissions Trading, Electricity Restructuring, and Investment in Pollution Abatement," *American Economic Review*, 100, 837–69.

GREENSTONE, M. AND I. NATH (2020): "Do renewable portfolio standards deliver cost-effective carbon abatement?" *University of Chicago, Becker Friedman Institute for Economics Working Paper (2019–62)*.

HORTAÇSU, A., S. A. MADANIZADEH, AND S. L. PULLER (2017): "Power to Choose? An Analysis of Consumer Inertia in the Residential Electricity Market," *American Economic Journal: Economic Policy*, 9, 192–226.

ISHII, J. AND J. YAN (2007): "Does divestiture crowd out new investment? The "make or buy" decision in the US electricity generation industry," *The RAND Journal of Economics*, 38, 185–213.

ITO, K. (2014): "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing," *American Economic Review*, 104, 537–563.

ITO, K. AND M. REGUANT (2016): "Sequential Markets, Market Power, and Arbitrage," *American Economic Review*, 106, 1921–57.

JHA, A. (2020): "Dynamic Regulatory Distortions: Coal Procurement at U.S. Power Plants," Available at SSRN 3330740.

JOSKOW, P. L. (1987): "Contract Duration and Relationship-Specific Investments: Empirical Evidence from Coal Markets," *The American Economic Review*, 77, 168–185.

——— (2005): "Regulation and Deregulation After 25 Years: Lessons Learned for Research in Industrial Organization," *Review of Industrial Organization*, 26, 169–193.

KAHN, A. E. (1988): "Surprises of Airline Deregulation," *The American Economic Review*, 78, 316–322.

KWOKA, J. (2008a): "Barriers to New Competition in Electricity Generation," *Report to the American Public Power Association, Northeastern University*.

——— (2008b): "Restructuring the U.S. Electric Power Sector: A Review of Recent Studies," *Review of Industrial Organization*, 32, 165–196.

MACKAY, A. (2022): "Contract duration and the costs of market transactions," *American Economic Journal: Microeconomics*, 14, 164–212.

MANSUR, E. T. (2007): "Upstream Competition and Vertical Integration in Electricity Markets," *The Journal of Law and Economics*, 50, 125–156.

MERCADAL, I. (2022): "Dynamic competition and arbitrage in electricity markets: The role of financial players," *American Economic Journal: Microeconomics*, 14, 665–99.

NEWBERY, D. M. AND M. G. POLLITT (1997): "The Restructuring and Privatisation of Britain's CEGB-Was It Worth It?" *The Journal of Industrial Economics*, 45, 269–303.

PULLER, S. L. (2007): "Pricing and Firm Conduct in California's Deregulated Electricity Market," *The Review of Economics and Statistics*, 89, 75–87.

RUBINOVITZ, R. N. (1993): "Market Power and Price Increases for Basic Cable Service Since Deregulation," *The RAND Journal of Economics*, 1–18.

RYAN, N. (2021): "The competitive effects of transmission infrastructure in the indian electricity market," *American Economic Journal: Microeconomics*, 13, 202–42.

SU, X. (2015): "Have Customers Benefited from Electricity Retail Competition?" *Journal of Regulatory Economics*, 47, 146–182.

VISCUSI, W. K., J. E. HARRINGTON JR, AND D. E. SAPPINGTON (2018): *Economics of Regulation and Antitrust*, MIT press.

Appendix

For Online Publication

A Details of Dataset Construction

In this section, we provide additional details about the construction of the dataset and state-specific deregulation.

A.1 Dataset Construction Details

Our dataset comes from several publicly-available data sources available from EIA and FERC. All data is reported annually. We construct our panel from 1994 through 2016.

Utility-level operational data were collected from form EIA-861. These data contain aggregate measures of generation, purchases, sales for resale, and retail sales for each utility. We combine these data with detailed retail and delivery sales (prices and quantities) by customer type, which is also from form EIA-861. We restrict our analysis to three types of customers: residential, commercial, and industrial, which account for the vast majority of retail consumption.[47] These data are reported at the utility-state level; for utilities that are located in multiple states, the combination of retail MWh and delivery MWh allows us to calculate each utility's total MWh serviced in each state. When constructing our data at the utility-state level, we scale aggregate variables from the operational data by the MWh serviced in each state (for multistate utilities only).

We obtained power plant generation data from forms EIA 759 between 1994 and 2000, EIA 906 between 2001 and 2007, and EIA 923 between 2008 and 2016. We used form EIA 906 for non-utilities generation during years 1999 and 2000. These data provide generator-specific measures of net generation and fuel consumption. For marginal costs, we use the average fuel cost of the upper quartile of MWh generated for all generators in a utility service area. We construct generator-specific and utility-specific marginal costs using the realized efficiency of each generator and the relevant fuel types. Unit fuel costs are estimated from purchased fuel receipts, which are reported in form EIA 423 for years prior to 2008 and form EIA 923 from 2008 onwards. When the unit cost of a given fuel was not available for a given power plant, we imputed it using the average unit cost for that fuel in the state and year. We obtain data on power plant operators from form 906, which we used to link each power plant to the utility

[47] The excluded customer types are transportation, public, and other.

that owned it pre-deregulation.[48] We use capacity data at the power plant level from EIA Form 860, which contains information on dates of initial operation and retirement.

Data on energy purchases were obtained from FERC Form 1. In this form, utilities report the identity of all sellers from which they purchased, as well as quantity, price, and other information. We identified whether each buyer-seller pair was affiliated via umbrella ownership under the same parent company by combining the information in a report on investor-owned utilities by the Edison Electric Institute (2019) and internet searches. We use the FERC Form 1 data to calculate the share of purchases from affiliated companies and the share of purchases from ISOs.

We manually constructed a panel of mergers and divestitures among the utilities in our dataset. We retroactively apply mergers to the entire panel and also undo divestitures, thus aggregating utilities that were ever part of the same entity into a single entity from the beginning to the end of the sample.

A.2 State-Specific Deregulation

To measure the impact of deregulation, we divide our sample into utilities in states that allowed for market-based electricity prices and those in states that continued with a state-sponsored monopoly and regulated rates. States that allowed for market-based electricity prices also enacted restructuring measures to allow for competitive entrants in the generation market (upstream) and in the retail market (downstream). Typically, incumbent utilities in deregulated states were no longer permitted to own generation facilities, but they were allowed to continue to operate downstream. Thus, retailers in deregulated states had to obtain electricity from a wholesale market, and consumers could choose between a regulated rate from the incumbent utility and market-based prices from independent retailers.

For each state, we identify whether deregulation measures were enacted, and when the measures legally came into effect. The 17 states that implemented deregulation measures in our period (1994–2016) are reported in Table A1, along the year of implementation. Upstream deregulation measures correspond to the vertical separation of a utility from generation facilities as well as an explicit allowance of competitive electricity suppliers. Downstream deregulation measures correspond to the introduction of a market for alternative retail suppliers. All of the states implemented these measures between 1998 and 2002, and the upstream and downstream legal changes typically occurred at the same time. Michigan is a notable exception, as they allowed for downstream competition but did not restructure the upstream market.

Five states—Arizona, Arkansas, Nevada, and Montana—initially passed deregulation measures but later rescinded them. We remove these four states from our analysis. We focus on investor-owned utilities (IOUs) that generated electricity in 1994. Because Nebraska and Ten-

[48] In the beginning of our sample, the operators coincided with ownership.

Table A1: First Year of Deregulation, by State

State	Implementation Year
NY	1998
RI	1998
CA	1999
NH	1999
MA	1999
ME	1999
CT	2000
DE	2000
MD	2000
NJ	2000
PA	2000
IL	2001
OH	2001
MI	2002
OR	2002
TX	2002
VA	2002

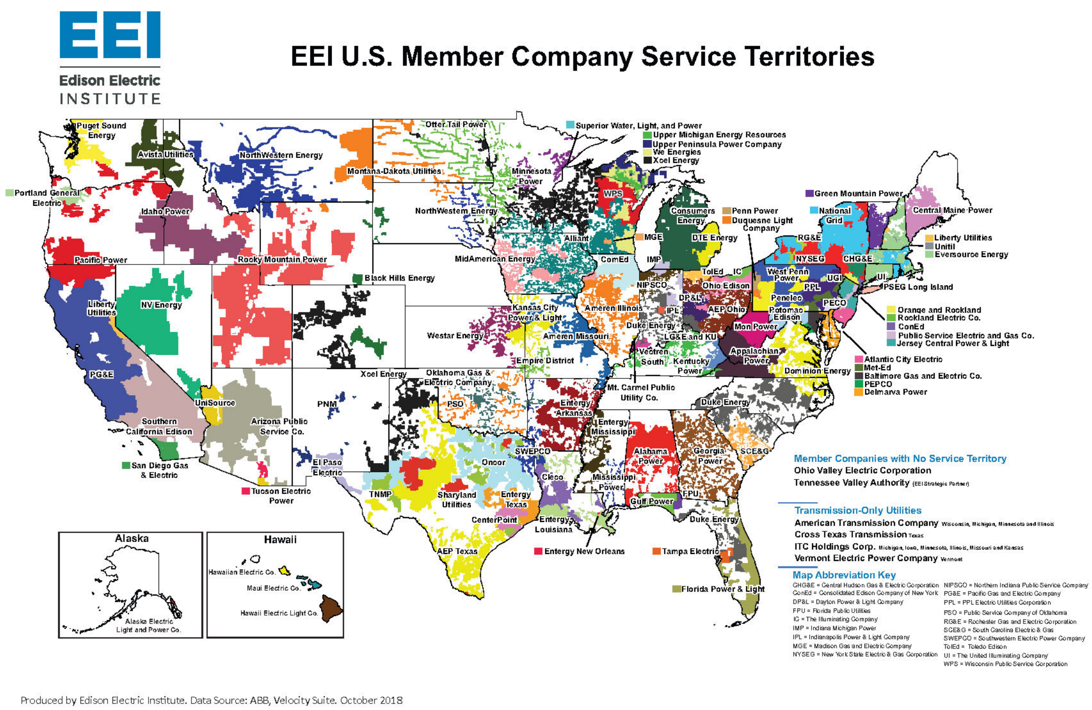
Notes: Table indicates the year initial deregulation measures came into effect for the listed states. For most states, this corresponds to when utilities began to divest generation assets. Michigan (MI) is an exception that did not pass a measure to deregulate the upstream market. Four states (AZ, AR, NV, and MT) initially passed deregulation measures but later rescinded them. These four states are omitted from our analysis.

nessee do not have utilities that meet these criteria, we also remove them from the analysis.[49] We are left with 17 states that introduced competitive markets and 25 states that did not. Our main sample consists of 78 treated utilities that were subject to deregulation measures and 75 utilities control utilities that were not.

Figure A1 presents a map of the geographic service areas for the utilities in United States. Our analysis focuses on the subset of these utilities that were in deregulated and control states that meet the above criteria.

[49] Nebraska does not have IOUs in this time period. In Tennessee, all generation comes from the federally operated Tennessee Valley Authority.

Figure A1: Areas Served by Investor-Owned Utilities



Notes: Figure displays the geographic service territories for investor-owned utilities in our sample as of 2018. Source: Edison Electric Institute.

B Conceptual Framework

Consider a regulator charged with implementing a regulated monopoly or market competition in an industry. The regulator seeks to maximize consumer welfare. Since consumer welfare is higher when consumers pay less, regulator chooses the design that minimizes consumer prices.

Let P_r denote regulated prices and P_m denote market-based prices. In the monopoly regime, prices are regulated to reimburse costs (c_r) and provide a regulated markup (μ_r) to reimburse the utility for its fixed-cost investments under rate-of-return regulation. The problem can be written as:

$$\min\{P_r, P_m\} \tag{4}$$

$$\text{s.t.} \quad P_r = c_r \cdot \mu_r. \tag{5}$$

Deregulation entails (i) the vertical separation of the upstream and downstream markets, (ii) entry by competitors in both markets, and (iii) prices determined by market forces. P_m may differ from P_r through different incentives to reduce costs and charge markups on marginal costs.

If the market is restructured, the utility is vertically separated and a wholesale market is created, such that the utility no longer generates its own electricity and now has to purchase it in the wholesale market at a price $w(c_m)$. This price will be a function of the marginal cost of production, c_m, which may be different from c_r because plants' operation, dispatch, and investment may change after restructuring:

$$P_m = w(c_m) \cdot \mu_m. \tag{6}$$

For simplicity, assume that $\mu_m = \mu_r$, i.e., the regulator does not change the permissible markup over procurement costs. By holding the retail markup fixed, we see from these two equations that the change in prices after deregulation depends on how the wholesale price w compares to the marginal cost under regulation c_r:

$$P_m < P_r \iff w(c_m) < c_r. \tag{7}$$

We can decompose this relationship into two components. The first reflects potential efficiency gains, which translate into lower costs under restructuring: $c_m < c_r$. The second component is the relationship between w and c_m, which depends on market power in the wholesale market. If the wholesale market is perfectly competitive, $w = c_m$. In this case, any efficiency gains resulting in $c_m < c_r$ will be passed on to prices, so $P_m < P_r$. Thus, a regulator anticipating perfectly competitive markets and efficiency gains would prefer market-based competition. This set of expectations rationalizes the widespread deregulation efforts observed in the U.S.

If the wholesale market is not perfectly competitive, upstream suppliers will charge a markup and the wholesale price will be

$$w = \frac{c_m}{1 + \frac{1}{\varepsilon}}, \tag{8}$$

where ε is the elasticity of the demand in the wholesale market. Suppliers will charge a positive markup as long as they face a demand that is less than perfectly elastic.

In addition, if the retail market is not perfectly competitive, retailers may be able to charge a markup μ_m that exceeds the regulated markup, μ_r. Thus, the presence of double marginalization—through larger retail margins ($\mu_m > \mu_r$) in addition to wholesale margins ($w - c_m$)—could outweigh the the efficiency gains that have been documented in the literature (Fabrizio et al., 2007; Cicala, 2015, 2022; Jha, 2020).

A regulator choosing between a regulated monopoly and market-based competition will choose deregulation if she expects efficiency gains to outweigh equilibrium markups, which depend on the degree of market power in the industry. If equilibrium markups are large relative to the efficiency gains, a regulated monopoly will ensure lower retail prices. This illustrates the important role of market power in designing regulations.[50]

[50] In this simple framework, which mirrors the discussion around deregulation in the U.S., the regulator's decision hinges on which regime provides the lowest prices. Regulators may also be concerned about elements outside of our framework, such as energy reliability and pollution. For example, pollution externalities could make higher prices more desirable from a welfare perspective. We believe that such considerations are better dealt with policies that target them directly (e.g., with taxes) rather than an inefficient pricing mechanism that may distort the market in other dimensions.

C Heterogeneity Across States

In the main text, we focus primarily aggregate effects across all states that implemented deregulation measures. Here, we examine the heterogeneity across states by calculating the average price effects using the utility-specific coefficients from our matching approach. To calculate the potential impact on consumer surplus, we assume an elasticity of -0.315, which is the estimated 5-year elasticity from Deryugina et al. (2019). We use the estimated price changes, the implied impact on quantities using the demand elasticity, and the realized values for prices and quantities to estimate the dollar impact on consumer surplus.

Table A2 reports the results. Panel (a) presents the annual averages over the full post-deregulation sample, from 2000 through 2016. On average, consumers paid 106 dollars per MWh for 1.4 petawatts of electricity in investor-owned utilities in deregulated states. We estimate an average price increase of 6.4 percent, corresponding to a decrease in quantity of 1.6 percent and an annual loss of $8.7 billion in consumer surplus.

We estimate some heterogeneity across states. Most deregulated states realized meaningful price increases, with 9 states realizing price effects exceeding 5 percent. We estimate that consumers in some states did benefit from deregulation, with consumers in Virginia and Illinois realizing meaningful decreases in prices.

As indicated by the earlier analysis, the effects increased over time. Panel (b) presents the annual results for the period 2006 to 2016, when we observe the realization of effective deregulation. We discuss timing in greater detail in the following section. From the later period, the estimated annual effects are greater, with aggregate price increases of 8.0 percent and annual loss in consumer surplus of $11.7 billion.

As a robustness check, we repeat the exercise using an alternative measure of price, which is an estimate of the realized retail prices using a within-state measure of delivery fees and retail prices from alternative suppliers. The results are reported in Table A3 in the Appendix. Overall, the results are similar but smaller in magnitudes, with annual losses in consumer surplus of $5.5 billion over the full sample and $7.5 billion from 2006 to 2016.

Table A2: Estimated Annual Impacts

(a) 2000–2016

State	Realized Values Price ($/MWh)	Realized Values Quantity (MWh)	Percent Change Price	Percent Change Quantity	Dollar Change Consumer Surplus
CA	138.91	189,195,740	21.4	-5.8	-4,426,780,852
NY	158.02	133,179,152	11.8	-2.9	-2,137,372,377
TX	88.43	250,568,361	9.1	-2.6	-1,627,322,560
CT	147.43	28,525,814	15.4	-3.9	-547,854,651
MD	104.57	58,544,944	8.9	-1.9	-516,473,654
MA	150.86	25,356,908	15.9	-4.3	-507,799,388
OH	83.89	135,971,028	5.6	-1.9	-499,096,124
OR	76.39	33,164,301	11.0	-3.2	-238,194,378
RI	134.22	7,624,773	17.7	-4.7	-149,456,983
NJ	123.81	72,405,815	1.4	-0.2	-122,194,363
NH	150.45	7,805,263	4.1	-0.8	-46,196,355
DE	99.28	8,657,489	1.7	-0.0	-18,724,126
ME	120.47	10,807,305	-0.9	0.3	13,456,753
MI	90.30	93,349,655	-1.6	0.7	129,487,139
VA	75.82	88,860,749	-4.0	1.3	272,619,229
PA	96.12	137,282,241	-3.5	1.2	477,672,144
IL	81.85	125,157,578	-14.5	4.9	1,728,510,030
All	106.30	1,406,457,117	6.1	-1.5	-8,215,720,517

(b) 2006–2016

State	Realized Values Price ($/MWh)	Realized Values Quantity (MWh)	Percent Change Price	Percent Change Quantity	Dollar Change Consumer Surplus
CA	146.41	194,617,260	24.2	-6.5	-5,284,272,352
NY	169.62	136,817,754	14.3	-3.5	-2,762,366,569
MD	125.62	57,191,583	18.8	-5.2	-1,097,300,888
TX	93.87	263,422,568	5.5	-1.5	-1,086,200,883
CT	172.10	28,144,316	26.1	-7.0	-958,223,951
OH	93.19	134,984,849	7.7	-2.6	-741,017,274
MA	167.80	25,618,216	21.1	-5.8	-711,361,414
NJ	137.55	73,251,746	5.7	-1.7	-512,230,406
OR	84.67	33,369,197	11.1	-3.3	-269,306,251
RI	148.98	7,665,444	25.2	-6.7	-220,400,082
NH	169.17	7,901,845	10.7	-2.8	-122,521,001
DE	118.60	8,391,985	9.1	-2.7	-81,249,387
MI	101.60	92,752,418	0.7	-0.2	-65,437,606
ME	128.66	11,042,416	-1.6	0.5	24,723,114
VA	82.66	92,361,718	-4.3	1.5	331,498,812
PA	104.26	141,719,115	-3.6	1.3	545,492,695
IL	88.25	128,082,698	-15.3	5.3	2,019,309,506
All	115.68	1,437,335,127	7.5	-1.8	-10,990,863,936

Notes: Impact on consumer surplus is calculated using the estimated price changes, the implied impact on quantities assuming a price elasticity of -0.315, and the realized values of prices and quantities. Price is the average bundled price weighted by the share of residential, industrial, and commercial customers served by each utility.

Table A3: Estimated Annual Impacts Using Alternative Price Measure

(a) 2000–2016

	Realized Values		Percent Change		Dollar Change
State	Price ($/MWh)	Quantity (MWh)	Price	Quantity	Consumer Surplus
CA	136.31	189,195,740	19.2	-5.3	-3,988,145,520
TX	88.43	250,568,361	9.1	-2.6	-1,627,322,070
NY	149.44	133,179,152	5.9	-1.5	-1,077,568,693
CT	146.29	28,525,814	14.5	-3.7	-519,400,405
MA	148.10	25,356,908	13.8	-3.8	-443,875,675
MD	100.57	58,544,944	4.9	-0.9	-294,981,097
OR	76.19	33,164,301	10.7	-3.1	-231,901,282
OH	81.25	135,971,028	2.4	-1.1	-161,121,391
RI	131.45	7,624,773	15.3	-4.1	-129,918,711
NJ	123.82	72,405,815	1.4	-0.3	-126,038,203
ME	120.47	10,807,305	-0.9	0.3	13,456,656
DE	94.39	8,657,489	-3.1	1.4	22,778,069
NH	136.13	7,805,263	-5.6	1.9	60,924,364
MI	89.32	93,349,655	-2.7	1.0	219,351,890
VA	75.82	88,860,749	-4.0	1.3	272,931,614
PA	91.66	137,282,241	-7.8	2.5	1,072,116,308
IL	80.33	125,157,578	-16.0	5.5	1,907,274,539
All	103.88	1,406,457,117	3.8	-0.9	-5,031,439,607

(b) 2006–2016

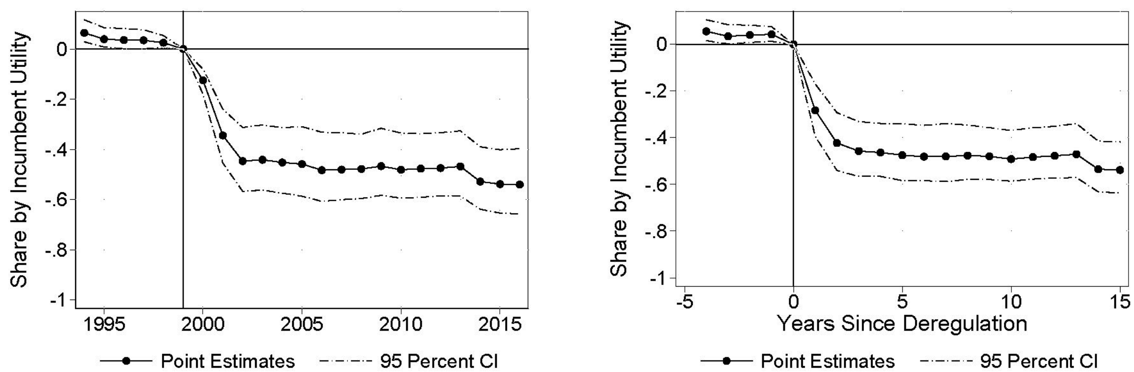
	Realized Values		Percent Change		Dollar Change
State	Price ($/MWh)	Quantity (MWh)	Price	Quantity	Consumer Surplus
CA	143.87	194,617,260	22.2	-6.0	-4,858,046,917
NY	159.72	136,817,754	7.7	-2.0	-1,516,114,355
TX	93.87	263,422,568	5.5	-1.5	-1,086,200,245
CT	170.34	28,144,316	24.9	-6.7	-915,221,756
MD	119.56	57,191,583	13.1	-3.8	-771,345,141
MA	164.49	25,618,216	18.7	-5.3	-635,588,424
NJ	137.43	73,251,746	5.6	-1.7	-508,440,217
OR	84.36	33,369,197	10.8	-3.2	-259,339,288
OH	88.96	134,984,849	2.9	-1.3	-202,687,791
RI	145.01	7,665,444	21.9	-6.0	-192,297,689
DE	111.00	8,391,985	2.1	-0.7	-18,634,737
ME	128.66	11,042,416	-1.6	0.5	24,723,001
NH	147.31	7,901,845	-3.6	1.2	43,029,004
MI	100.12	92,752,418	-0.7	0.3	68,606,369
VA	82.65	92,361,718	-4.3	1.5	332,122,354
PA	97.61	141,719,115	-9.6	3.2	1,458,145,195
IL	86.12	128,082,698	-17.3	6.0	2,275,019,469
All	112.52	1,437,335,127	4.7	-1.1	-6,762,271,169

Notes: Impact on consumer surplus is calculated using the estimated price changes, the implied impact on quantities assuming a price elasticity of -0.315, and the realized values of prices and quantities. Price is a measure of the average price paid by all retail consumers in each utility's service area. It is calculated as the weighted average of the bundled service price and an approximate measure of the retail price to customers of alternative electric suppliers. The approximate measure is constructed as the sum of utility-specific delivery services and the statewide average retail energy price. We do not have utility-specific measures of energy prices from alternative suppliers, and reporting may vary across utilities due to lack of standardization.

D Robustness Checks and Additional Analysis

Comparison of Event Timing Approaches

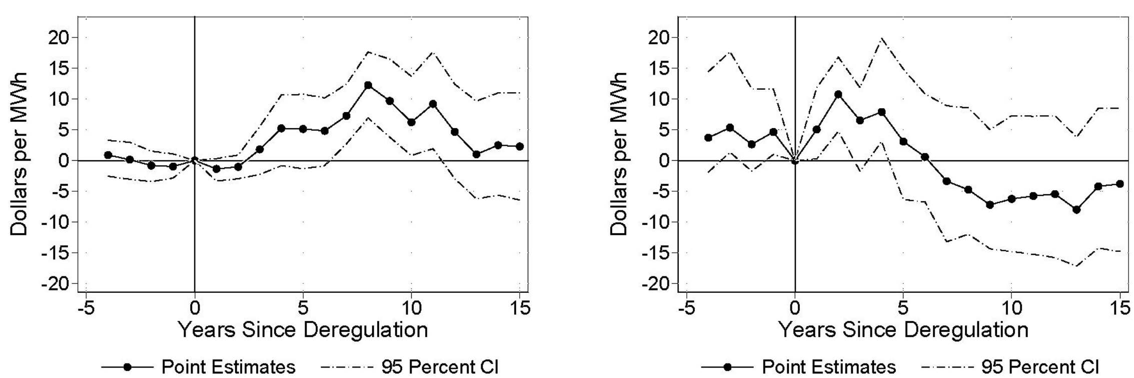
Figure A2: Different Choices of Timing



(a) Calendar Time (b) Utility-Specific Event Time

Notes: Figure displays difference-in-differences matching estimates of changes to incumbent utilities share of quantity demanded provide by its own generation. Panel (a) displays the results in calendar years, following the results in the main text. Panel (b) displays the results indexed to time period 0, which represents the year prior to the implementation of deregulation measures in each utility's state. The dashed lines indicate 95 confidence intervals, which are constructed via subsampling.

Figure A3: Event Study Estimates of Changes in Prices and Costs After Deregulation

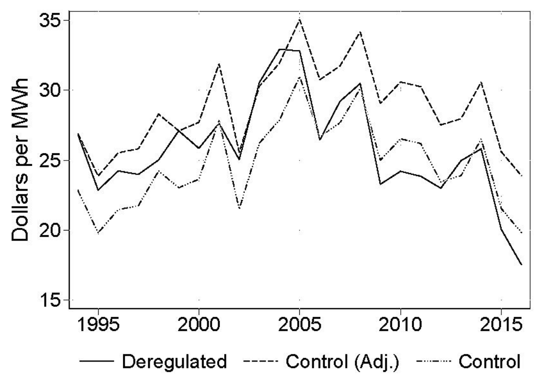


(a) Retail Prices (b) Generation Fuel Costs

Notes: Figure displays difference-in-differences matching estimates of changes in (a) retail prices and (b) fuel costs for deregulated utilities. Each deregulated utility is matched to a set of three control utilities based on 1994 characteristics. The estimated effects are indexed to time period 0, which represents the year prior to the implementation of deregulation measures in each utility's state. The dashed lines indicate 95 confidence intervals, which are constructed via subsampling.

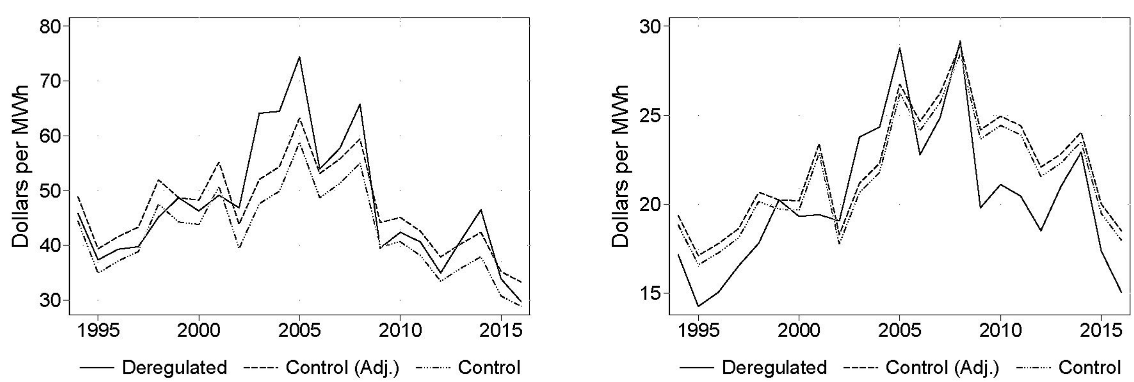
Alternative Measures of Generation Costs

Figure A4: Average Variable Fuel Costs



Notes: Figure plots the average fuel costs of generation for all generating facilities in deregulated states (solid black line) and control states (grey line). The dashed line plots retail prices and fuel costs for control states after adjusting for level differences in 1999.

Figure A5: Statewide Fuel Costs



(a) Marginal Costs (b) Average Variable Costs

Notes: Figure plots the statewide measure fuel costs using our measure of marginal costs and average variable costs. Marginal costs are calculated as the average fuel costs for the 75th percentile and up of MWh generated for all generating facilities in deregulated states (solid black line) and control states (grey line). The dashed line plots retail prices and fuel costs for control states after adjusting for level differences in 1999.

Difference-in-Differences Effects with Average Variable Costs

Table A4: Relative Changes in Prices, Costs, and Markups (AVC)

	(1)	(2)	(3)	(4)	(5)	(6)
	Retail Price	Wholesale Price	Generation Cost (AVC)	Retail Markup	Wholesale Markup	Gross Markup
1999 Values	78.06	42.81	26.61	34.95	17.09	51.40
2000-2005	4.14	−0.42	−2.75	4.87	2.29	6.85
	(1.74)	(2.97)	(2.83)	(2.45)	(4.66)	(3.64)
2006-2011	12.73	3.46	−9.12	9.38	11.60	21.71
	(2.95)	(3.49)	(3.32)	(3.84)	(4.97)	(4.60)
2012-2016	5.83	7.41	−8.63	2.63	14.88	14.60
	(3.83)	(4.18)	(3.32)	(4.10)	(5.39)	(4.98)
2000-2016	7.66	3.16	−6.71	5.62	9.12	14.31
	(2.30)	(2.99)	(2.65)	(2.73)	(4.16)	(3.74)

Notes: Table displays the estimated difference-in-differences matching cofficients for prices, costs, and markups between deregulated and control utilities in dollars per MWh. In this table, costs and markups are calculated using average variable costs. The first row provides the baseline values in 1999, and the remaining rows provide the average effect for the specified time period. Standard errors are displayed in parentheses.

Matching with Geographic Proximity

Here, we report summary statistics (Table A5) and difference-in-differences results (Table A6) when we also match on Census region. The results are very similar to the baseline specification.

Table A5: Characteristics of Deregulated and Alternative Matched Control Utilities in 1994

	(1)	(2)	(3)	(4)	(5)
	Deregulated	Control		Matched Controls	
	Mean	Mean	p-value of Difference from (1)	Mean	p-value of Difference from (1)
ln(MWh Retail)	15.21	15.22	0.977	15.40	0.717
ln(MWh Generated)	14.70	14.60	0.857	14.59	0.891
Marginal Generation Share: Coal	0.50	0.54	0.705	0.53	0.817
Marginal Generation Share: Gas	0.12	0.15	0.639	0.12	0.943
Marginal Generation Share: Nuclear	0.02	0.02	0.763	0.01	0.575
Marginal Generation Share: Oil	0.19	0.07	0.078	0.16	0.735
Marginal Generation Share: Water	0.18	0.20	0.763	0.18	0.960
Marginal Fuel Costs	65.69	37.89	0.137	59.11	0.795
Retail Price	78.76	58.95	0.001	59.78	0.002
Number of Unique Utilities	78	76		72	

Notes: Table displays 1994 characteristics for 78 investor-owned utilities in states that later deregulated and 76 investor-owned utilities in states that did not deregulate. Columns (1) and (2) report the mean characteristics for each group, and column (3) reports the p-value of the difference in means. Column (4) reports the means for matched controls using a nearest-neighbor methodology, and column (5) reports the p-value of the difference in means between matched controls and the deregulated utilities. The first eight variables are used as matching variables, along with Census region.

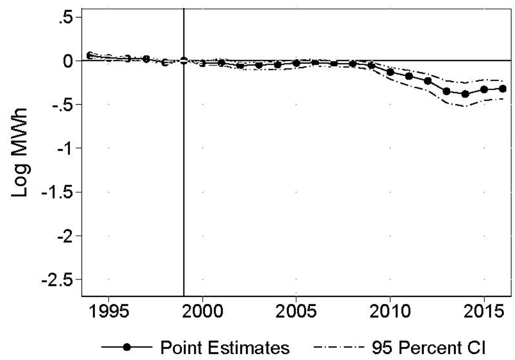
Table A6: Relative Changes in Prices, Costs, and Markups (Geographic Matching)

	(1)	(2)	(3)	(4)	(5)	(6)
	Retail Price	Wholesale Price	Generation Cost	Retail Markup	Wholesale Markup	Gross Markup
1999 Values	78.06	42.81	48.89	34.95	−5.22	29.13
2000-2005	4.14	−0.42	−0.63	4.87	−0.26	4.74
	(1.74)	(2.97)	(2.88)	(2.45)	(4.52)	(3.59)
2006-2011	12.73	3.46	−10.59	9.38	12.30	23.18
	(2.95)	(3.49)	(4.82)	(3.84)	(6.03)	(5.72)
2012-2016	5.83	7.41	−10.83	2.63	16.40	16.80
	(3.83)	(4.18)	(4.92)	(4.10)	(6.56)	(6.01)
2000-2016	7.66	3.16	−7.11	5.62	8.82	14.71
	(2.30)	(2.99)	(3.59)	(2.73)	(4.68)	(4.40)

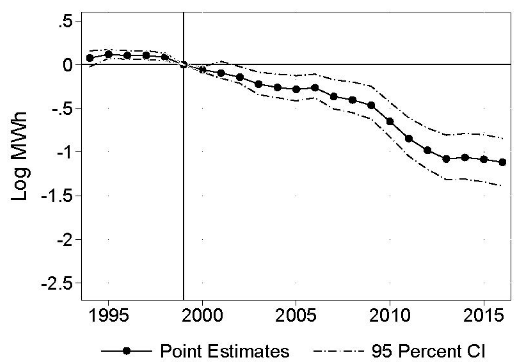
Notes: Table displays the estimated difference-in-differences matching cofficients for prices, costs, and markups between deregulated and control utilities in dollars per MWh. The first row provides the baseline values in 1999, and the remaining rows provide the average effect for the specified time period. Standard errors are displayed in parentheses. The results correspond to a specification with geographic proximity as a matching variable.

Change in Downstream Consumption

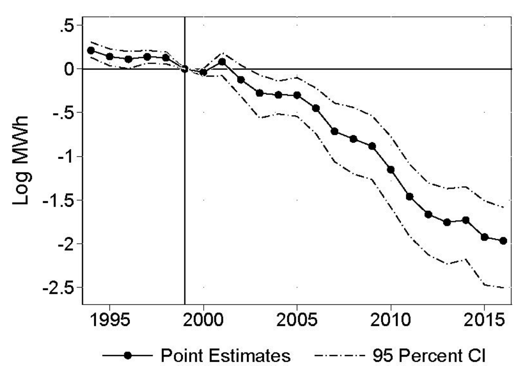
Figure A6: Change in Incumbent Utility Retail MWh (Bundled Service)



(a) Residential



(b) Commercial

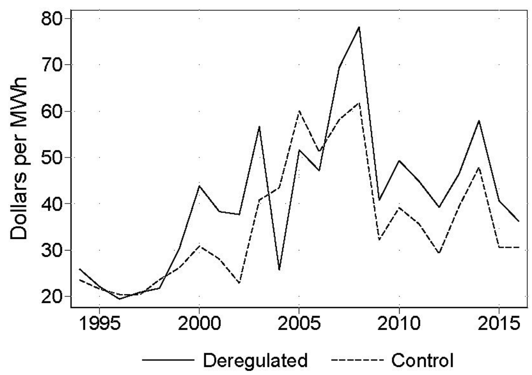


(c) Industrial

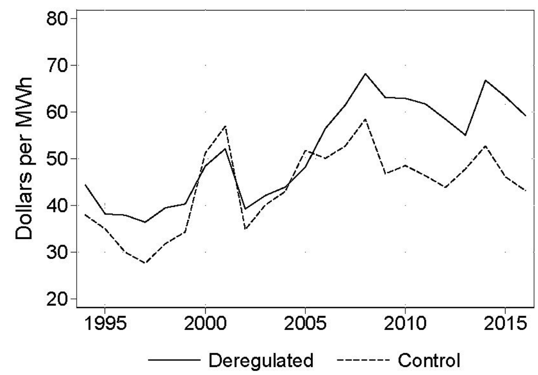
Notes: Figure displays difference-in-differences matching estimates of changes in log MWh for bundled service for deregulated utilities. Bundled service customers are those remaining on regulated rates in deregulated areas. We exclude Texas and Maine, which fully eliminated bundled service. Each deregulated utility is matched to a set of three control utilities based on 1994 characteristics. The estimated effects are indexed to 1999, which is the year prior to the first substantial deregulation measures. The dashed lines indicate 95 confidence intervals, which are constructed via subsampling.

Wholesale Electricity Markets: ISOs and Bilateral Contracts

Figure A7: Wholesale Prices from Spot Markets (ISOs and Power Pools) and Bilateral Contracts



(a) Spot Markets



(b) Bilateral Contracts

Notes: Figure displays the wholesale prices based on utility-level purchases for deregulated states (solid lines) and control states (dashed lines). Panel (a) plots the MWh-weighted average purchase prices from ISO markets and power pools, and panel (b) plots the MWh-weighted average from bilateral contracts.

MACKENZIE

ORANGE

B  K

2023 LONG-TERM CAPITAL MARKETS OUTLOOK

ORANGE BOOK

2023 LONG-TERM CAPITAL MARKETS OUTLOOK

Contents

Capital market assumptions	**3**
Long-term expected returns	4
Volatility and correlations	7
Currency valuations	8
Methodology	9
Asset allocation	**10**
Pension plans and funding risk	11
Canadian pension landscape	14
Currency hedging and overlays	16
Macroeconomic factors	21

Introduction

Todd Mattina, PhD
Co-Lead, Multi-Asset Strategies Team

Nelson Arruda, MFin., MSc., CFA
Co-Lead, Multi-Asset Strategies Team

Jules Boudreau, MA
Economist, Multi-Asset Strategies Team

Mackenzie presents the 2023 Orange Book, our long-term outlook on domestic and global markets. Here, we highlight our expectations for the average return of stocks and bonds over the coming decade.

Day-to-day moves in financial markets make the headlines. But what really matters for long-term investors is their total portfolio return over longer investment horizons. The return estimates presented in the Orange Book help investors look through short-term market movements to stay focused on the long term.

Our capital market assumptions are also appropriate for sophisticated institutional investors, such as pension funds and endowments. Long-term risk and return expectations are key inputs for strategic allocations.

In the second section, starting on page 10, we cover four topics relevant to institutional investors: funding risk management, fund allocation, currency hedging and macro risk management.

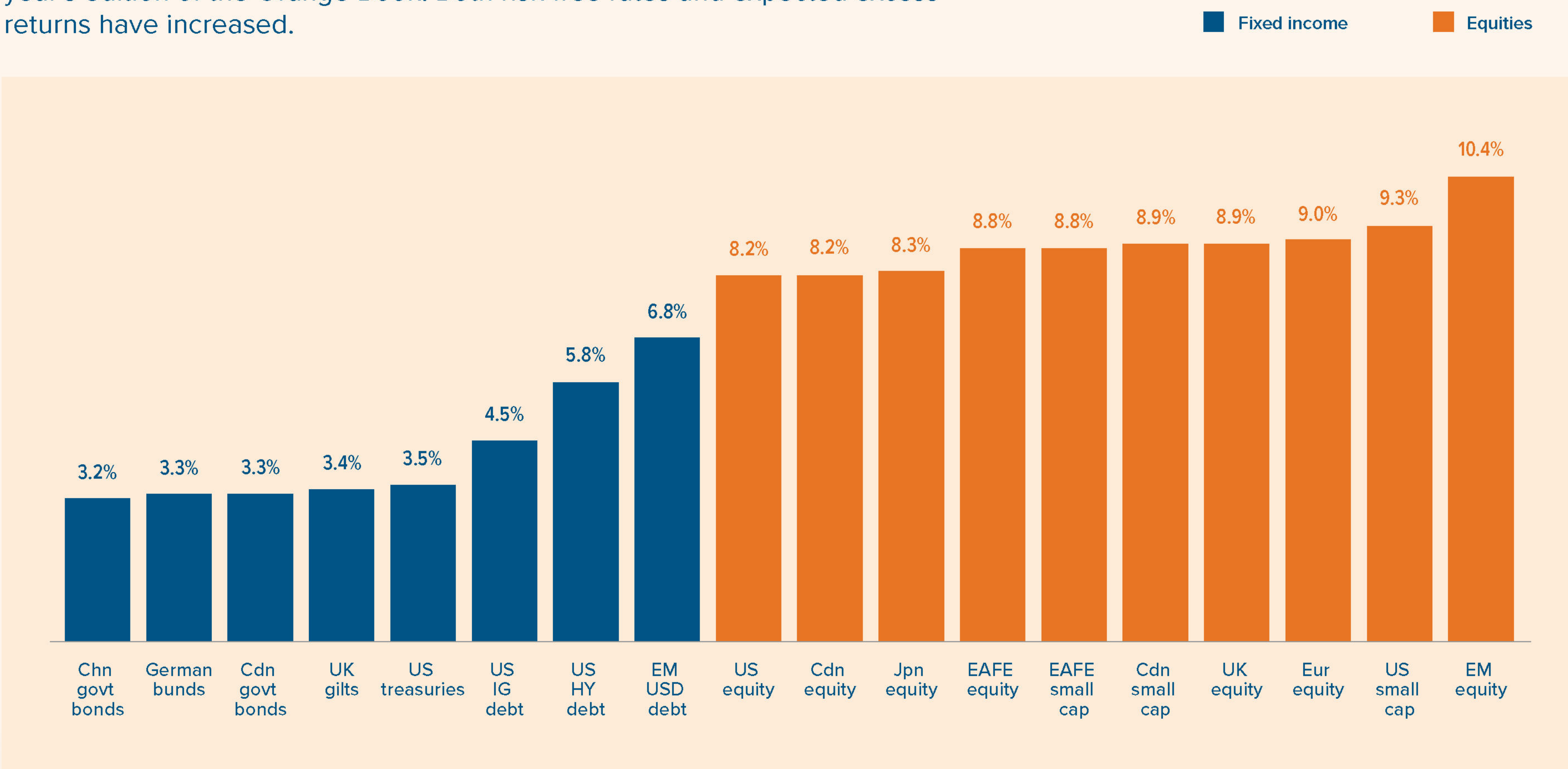
ORANGE BOOK
2023 LONG-TERM CAPITAL MARKETS OUTLOOK

Capital market assumptions

- 10-year expected returns
- 5-year expected returns
- 10-year expected returns vs. risk
- Expected asset class volatility and correlations
- Currency valuations
- How we estimate expected returns

10-year expected returns (FX hedged)

Long-term expected returns have risen for all assets in our universe from last year's edition of the Orange Book. Both risk-free rates and expected excess returns have increased.



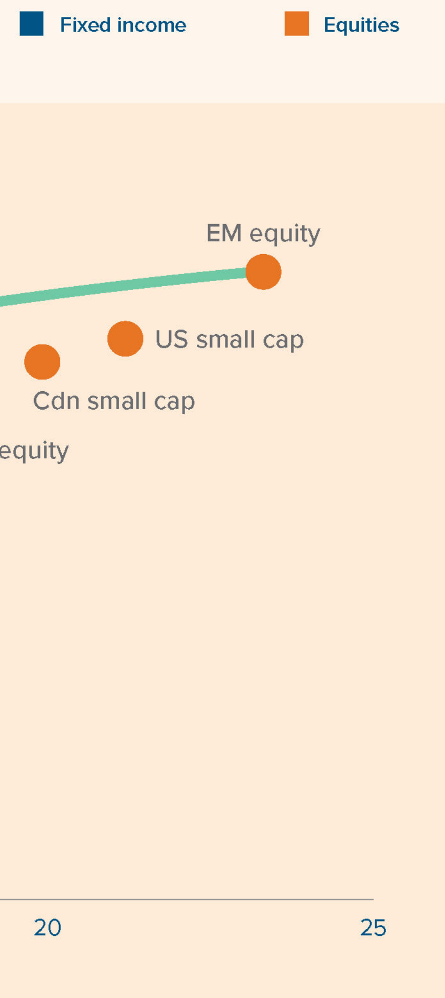
Expected geometric returns are shown on a nominal basis, before fees for all asset classes. Please refer to the following page for our five-year expected annual returns, where the active expected return component based on our value, macro and sentiment models play a greater role in shaping expected returns. Developed-market sovereign bond returns shown here reflect the expected return to investing in a constant-maturity 10-year government bond. Estimated using data as of November 30, 2022.

5-year expected returns (FX hedged)

Over a five-year horizon, expected returns are driven significantly by starting asset valuations and economic conditions. We expect relatively expensive US stocks to underperform other stock markets.

Expected geometric returns are shown on a nominal basis, before fees for all asset classes. The five-year return expectations have a greater weight in our own active views, which will have more weight over a five-year horizon than over 10 years. Estimated using data as of November 30, 2022.

10-year expected returns vs. risk



Expected geometric returns are shown on a nominal basis, before fees for all asset classes. These are contrasted with each asset's expected monthly annualized volatility. Estimated using data as of November 30, 2022

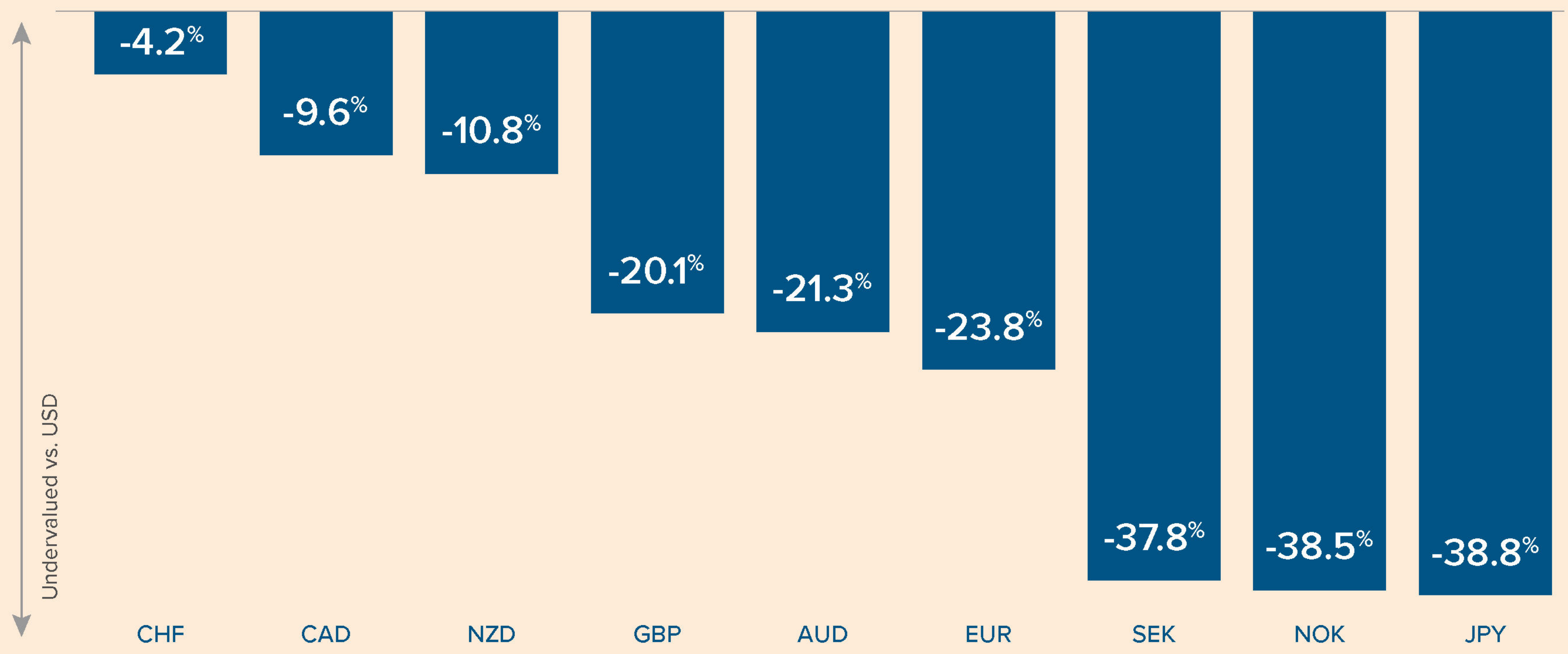
Expected asset class volatility and correlations

	Volatility	US treasuries	Cdn govt bonds	German bunds	UK gilts	US IG debt	US HY debt	EM USD debt	Chn govt bonds	US equity	Cdn equity	Jpn equity	UK equity	Eur equity	EM equity	US small cap	Cdn small cap	EAFE small cap	EAFE equity
US treasuries	6.0%	1.0																	
Cdn govt bonds	5.8%	0.8	1.0																
German bunds	5.4%	0.7	0.7	1.0															
UK gilts	6.2%	0.7	0.7	0.8	1.0														
US IG debt	6.5%	0.7	0.6	0.6	0.5	1.0													
US HY debt	9.7%	0.0	0.1	0.1	0.1	0.6	1.0												
EM USD debt	12.4%	0.3	0.3	0.3	0.3	0.6	0.7	1.0											
Chn govt bonds	4.6%	0.1	0.2	0.2	0.2	0.0	-0.2	-0.2	1.0										
US equity	17.5%	0.0	0.0	0.0	0.1	0.4	0.7	0.6	-0.1	1.0									
Cdn equity	14.9%	-0.1	0.0	0.0	0.0	0.4	0.6	0.6	-0.2	0.8	1.0								
Jpn equity	18.3%	-0.2	-0.2	-0.2	-0.1	0.2	0.5	0.4	-0.3	0.6	0.5	1.0							
UK equity	15.2%	-0.1	0.0	0.0	0.1	0.3	0.5	0.5	0.0	0.7	0.7	0.5	1.0						
Eur equity	17.2%	-0.1	-0.1	0.0	0.0	0.3	0.6	0.5	-0.1	0.8	0.7	0.6	0.8	1.0					
EM equity	23.3%	0.0	0.0	-0.1	0.0	0.4	0.6	0.7	-0.2	0.7	0.7	0.6	0.6	0.6	1.0				
US small cap	21.2%	-0.1	-0.1	0.0	0.0	0.3	0.7	0.5	-0.1	0.9	0.8	0.6	0.6	0.7	0.7	1.0			
Cdn small cap	19.9%	-0.1	0.0	0.0	0.0	0.4	0.6	0.6	-0.2	0.7	0.8	0.5	0.6	0.5	0.7	0.7	1.0		
EAFE small cap	15.6%	-0.2	0.0	-0.1	0.0	0.3	0.6	0.5	-0.2	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7	1.0	
EAFE equity	13.2%	-0.2	-0.1	-0.1	0.0	0.3	0.6	0.6	-0.2	0.8	0.7	0.8	0.8	0.9	0.7	0.8	0.6	0.9	1.0

Expected monthly annualized volatility and monthly returns correlations. Estimates are based on exponential decay-weighted monthly returns over the 1900-2022 period, adjusted for an unbalanced sample.

Currency valuations

Among G10 currencies, the US dollar is the most overvalued, while the Japanese yen is the cheapest relative to long-term fair value. We expect the US dollar to depreciate against all currencies over the coming decade.



These measures of over- and undervaluation incorporate four of our assessments of long-term and medium-term currency valuation. We assess valuations based on a proxy for absolute purchasing power parity, real effective exchange rates, a behavioural terms-of-trade adjusted currency valuation model, and another behavioural model that adjusts balance-of-payments outcomes based on structural economic factors. Estimated using data as of November 30, 2022.

How we estimate expected returns

Long-term expected asset return = Excess returns + Risk-free rates

Excess returns compensate investors for bearing risk and can vary as investors' risk appetite fluctuates with economic and financial conditions.

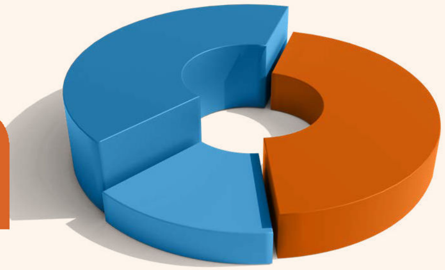
Risk-free rates are determined from the current yield curve and reflect the central bank's policy interest rate, expected inflation and growth.

Excess returns = Risk premiums + Expected active returns

Risk premiums represent a systematic source of excess return linked to the asset class volatility and its correlation to the global capital market portfolio.

Expected active returns are expected shifts in the asset return from its long-term risk premium. Expected active returns reflect proprietary insights about valuation, macro conditions and investor sentiment.

Asset allocation



- Pension plans and funding risk
- Canadian pension landscape
- Currency hedging and overlays
- Macroeconomic factors

Pension plans and funding risk

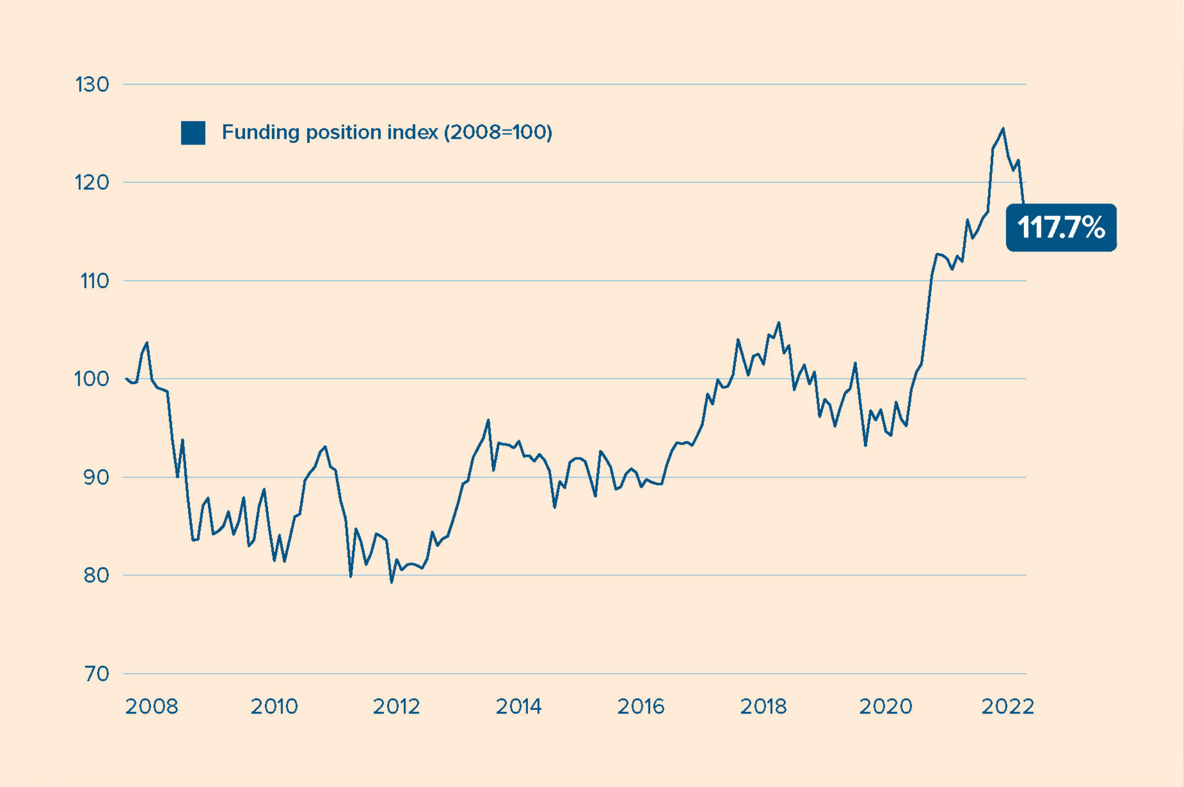
Improved funding position for the average Canadian pension plan

Pension plans had a surprisingly positive year in 2022, despite a challenging period in capital markets. While asset prices generally fell, the dramatic rise in long-term interest rates reduced the present value of future pension liabilities by even more than the fall in assets (see chart). With funding ratios (solvency basis) exceeding 100% for many pension funds, plan sponsors will soon need to decide how to manage surplus positions.[1] They have three main options:

- De-risking their asset allocation.
- Using their plan surplus to enhance benefits.
- Enhancing the asset mix to reduce surplus risk and improve risk-adjusted expected returns.

Each option has merits and pitfalls. Overall, we see a compelling case for plans to enhance their asset mix, reduce surplus risk to “lock-in” improved long-term funding positions, and leave critical plan parameters unchanged, such as the inflation indexation of benefits, given the uncertain economic environment in 2023 and 2024.

Canadian pension plans: Model-based index of funding position[2]



1 According to FSRA, the median funding ratio in Ontario is about 109% with 78% of plans reporting a solvency ratio above 100%. Mercer reports similar findings in its database of pension clients nationally.
2 Calculations for the funding risk index by the Mackenzie Multi-Asset Strategies Team, using Canadian wage growth data via the Bank of Canada, duration-adjusted corporate spreads via Bloomberg, and asset mix data via the Pension Investment Association of Canada. Based on a solvency basis approach.

1. De-risking asset allocation:

For sponsors in a comfortably fully funded position, de-risking the plan may be an attractive option. De-risking can take different shapes. One approach is to transfer the plan's liabilities and assets via a "pension risk transfer" (PRT).[1] For fully funded plans, PRT can reduce the sponsor's risk of unexpected special payments and better align changes in asset values with changes in future pension liabilities. The PRT market has expanded in recent years with pricing dependant on multiple factors, including composition of assets (i.e., equity, high yield bonds, etc.) and assumed longevity of the pensioners.

For plans with internal management capabilities, expanding interest rate sensitivity on the asset side is an alternative approach to de-risk the portfolio. In this way, the portfolio can be customized to match the inflation and interest rate sensitivity of the plan's own pension liabilities. Expanding interest rate sensitivity typically involves the use of both leverage and derivatives, such as interest rate swaps, so it is critical for plans to manage liquidity effectively. As made clear by the aftermath of the UK mini-budget debacle, a sudden surge in interest rates can require pension funds to raise liquidity abruptly to cover losses in leveraged positions. A key lesson is that liability-aware pension strategies can reduce long-term surplus risk but also increase short-term liquidity risk. De-risking effectively requires that plans balance this trade-off effectively.

Liability-aware pension strategies can reduce long-term surplus risk but also increase short-term liquidity risk

2. Enhancing pension benefits:

Given the breakout in inflation in 2022, sponsors could also face pressure to enhance pension benefits, including by increasing inflation indexation. A comfortable surplus position could provide room to enhance benefit policies. However, as elaborated below, long-term funding positions can reverse quickly with a change in economic conditions. For instance, an unexpected hard landing in the economy next year could both lower long-term interest rates and depress asset values, reversing recent gains in funding positions. Inflation could also be stickier than expected, raising the long-term cost of inflation indexation provisions. Consequently, we believe that plan sponsors should maintain a modest-to-moderate surplus position as a precautionary buffer in 2023.

An unexpected hard landing in the economy next year could both lower long-term interest rates and depress asset values

1 Pension Risk Transfer in Canada and the US, B. Simmons, SOA Research Institute, February 2022.

3. Enhancing asset allocation:

While long-term funding positions have improved this year, many pension plans are one recession away from renewed challenges. In a typical hard landing for the economy, equities and other risk assets decline in value as investors require wider risk premiums, and long-term interest rates fall as investors demand the safety of government bonds. Lower long-term rates imply a higher present value of future pension liabilities, just as asset valuations are falling. A surplus position can evaporate quickly in this scenario.

To monitor and control this risk, sponsors should evaluate the sensitivity of long-term funding ratios to a hard landing in 2023 when stress-testing alternative scenarios. Many economists expect a short and shallow recession in 2023 as the base case, however a hard landing with high and sticky inflation remains a feasible alternative scenario in our view. In advanced economies, the historical track record of reducing inflation from high levels suggests that it could take several years to bring inflation down from high single-digit rates to 2%.[1] A prolonged period of high interest rates could be needed to cool labour markets and prevent a wage-price spiral. Notable economists argue that the US unemployment rate may need to rise from a near record low of 3.7% in late 2022 to over 5% to quell inflationary pressures.[2]

Enhancements to a plan's asset mix can set the stage for a more durable improvement in long-term funding positions. Potential enhancements include:

Adding **interest rate sensitivity** on the asset side to better match the factors driving changes in liabilities, reducing the plan's surplus risk.

- Adopting modest leverage allows for greater interest rate sensitivity without sacrificing market exposure to return-seeking asset classes, such as equities.

Reducing **risk concentrations** on the asset side, such as "home bias" in equity allocations and under-allocations in international equities (i.e., UK, Europe, Japan and EM stocks) relative to country weights based on market capitalization (see p. 15).

Expanding allocation to **alternative assets and investment strategies** to broaden the range of return drivers in the portfolio, expand the opportunities to add value and adopt risk-diversifying strategies that can compete with equities.

Enhancing **FX management** to reduce total portfolio risk (see p. 17).

- Maintain long USD exposure to balance foreign equity risk.
- Hedge pro-cyclical and commodity currencies that are correlated to CAD.

Balancing **long-term funding risk with short-term liquidity risk** — avoid suffering the same fate as UK pension plans with LDI strategies.

- For private assets, smoothing is a key advantage for sponsors at risk of special payments if funding ratios decline.
- Leverage, FX management and liquid alt strategies require use of derivatives that require cautious liquidity management.

1 "History Lessons: How 'Transitory' Is Inflation", R. Arnott, November 2022. https://www.researchaffiliates.com/publications/articles/965-history-lessons
2 See L. Summers, June 20, 2022, Bloomberg. https://www.bnnbloomberg.ca/larry-summers-says-us-needs-5-jobless-rate-for-five-years-to-ease-inflation-1.1781433.

Canadian pension landscape

Pension plans face three key risks in funding long-term liabilities:

1. Short duration, because of a mismatch between the risk factors driving asset returns and liability growth.
2. Concentrated equity risk on the asset side.
3. Currency risk.

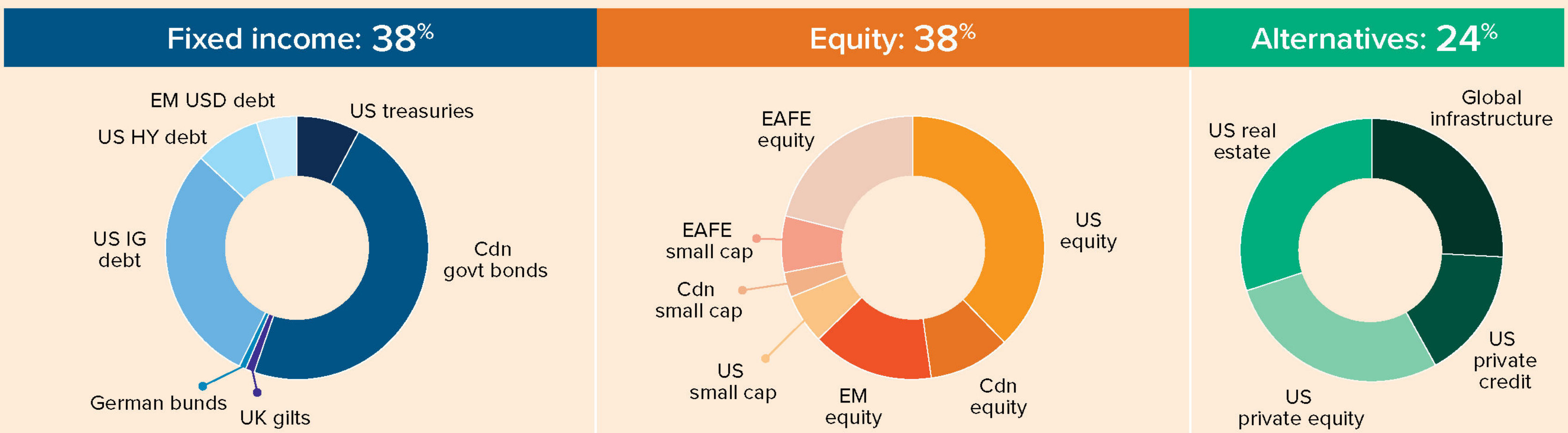
Pension plans have transitioned their strategic asset allocations to manage these risks:

1. Higher allocations to alternative assets.
2. Lower allocations to public market equities.
3. Greater diversification within asset class categories.
4. Leverage to increase interest rate sensitivity, balancing portfolio exposures while improving risk-adjusted expected returns and asset-liability surplus risk.

The average DB plan also maintains a significant allocation to liquid fixed income securities.

1. A liquidity buffer provides room for covering capital calls from private asset managers, FX hedges and rebalancing.
2. However, modest leverage limits room to extend interest rate sensitivity.

Average portfolio weights



Benchmark pension portfolio, shown in capital space, constructed by the Mackenzie Multi-Asset Strategies Team using our universe of asset classes, based on the Pension Investment Association of Canada's (PIAC) report on average Canadian pension plan holdings.

60/40 vs. PIAC average

The average Canadian pension fund's allocation deviates significantly from a vanilla 60/40 portfolio. The average pension fund allocates 25% to alternatives, retains a sizable "home bias" in Canadian equities, underweights EM equities and maintains a smaller allocation to bond duration.

The average pension portfolio has a higher expected return than a 60/40 and lower volatility. The pension portfolio is also expected to exhibit a lower surplus risk, i.e., lower volatility in the difference between the value of pension assets and liabilities.

The average pension fund adopts modest leverage. Increasing leverage could allow pension funds to better match their assets to their liabilities and to manage FX exposures efficiently. But in choosing a leverage ratio, funds must balance long-term funding risk with short-term liquidity constraints (see p. 13).

For private asset classes, the observed volatility will tend to be lower than the "true" mark-to-market volatility. Given that many alternatives are only valued periodically, observed volatility is artificially smoothed compared to public market assets. Because pension sponsors should care about both the observed volatility and the mark-to-market volatility of their portfolio, we use a 50/50 blend of observed and modeled volatilities for our private assets' risk estimates.

Asset class	60/40	PIAC average
US treasuries	0.0%	3.0%
Cdn govt bonds	22.4%	19.3%
German bunds	0.0%	0.3%
UK gilts	0.0%	0.5%
IG debt	16.8%	12.0%
HY debt	0.8%	3.1%
EM USD debt	0.0%	1.9%
US equity	25.4%	15.4%
Cdn equity	1.8%	4.0%
Jpn equity	3.5%	2.2%
UK equity	1.9%	1.2%
Eur equity	7.9%	5.0%
EM equity	13.8%	5.9%
US small cap	2.4%	2.5%
Cdn small cap	0.2%	1.2%
EAFE small cap	3.0%	2.7%
Global infrastructure	0.0%	6.5%
US private credit	0.0%	3.9%
US private equity	0.0%	7.0%
US real estate	0.0%	7.4%
Proportion fixed income	40.0%	40.1%
Proportion equity	60.0%	40.0%
Proportion alts	0.0%	24.8%
Expected return (10-year average)	7.0%	7.5%
Volatility	10.7%	9.4%
Sharpe	0.38	0.48
Surplus risk	11.8%	10.1%
Tracking error vs. 60/40	–	1.9%
Total exposure (including leverage)	100.0%	104.8%

Calculations by the Mackenzie Multi-Asset Strategies team based on our estimates of expected returns, volatilities, and correlations. Benchmark pension portfolio, shown in exposure space, constructed by the Mackenzie Multi-Asset Strategies Team using our universe of asset classes, based on the Pension Investment Association of Canada's (PIAC) report on average Canadian pension plan holdings, making reasonable assumptions as to the decomposition of global holdings. Asset returns are shown gross of fees, including for alternative assets, which typically exhibit high fees. For private asset classes, we use a 50/50 blend of observed (smoothed) and modeled (de-smoothed) volatilities.

Currency hedging and overlays

Fully hedging currency risk in portfolios with foreign assets is rarely optimal for risk minimization. For example, exposure to reserve currencies, such as the US dollar, can reduce total portfolio risk in local currency terms.

Investors can exploit the correlation of a currency with foreign asset returns and other currencies to identify an optimal FX hedge ratio based on risk minimization of the total portfolio. Optimal currency hedge ratios will depend on an investor's investment horizon (p. 18), their home currency, their risk aversion, and the composition of their portfolio.

Going a step further, investors can dynamically hedge their FX exposures to take advantage of time-varying expected returns for currencies. Historically, active investment strategies in the currency space have generated excess returns, providing a diversifying source of value-add in the portfolio. In particular, models based on relative valuation, macroeconomic factors and investor sentiment have a good track record at delivering risk-adjusted active returns.

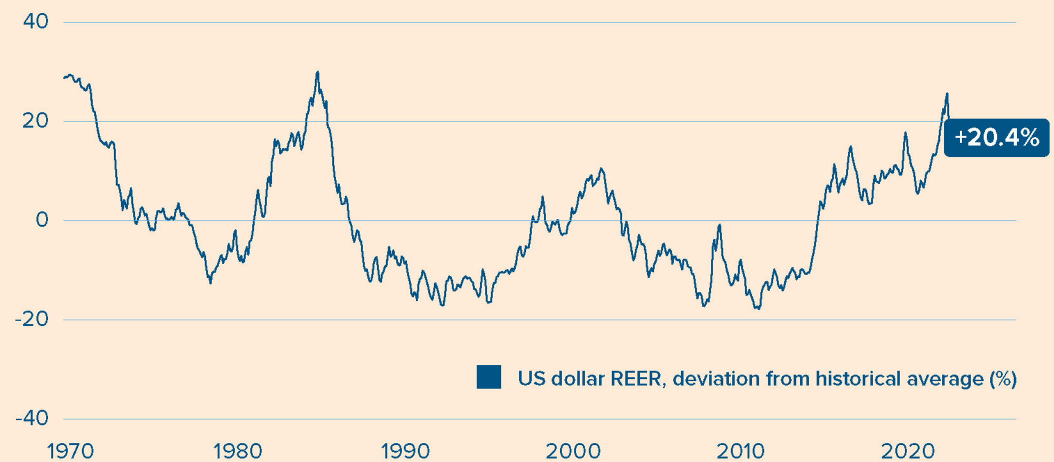
Currency management can be outsourced to specialized managers using a portable overlay strategy. In this way, FX management goes from a problem to solve to an independent source of expected active return. By using derivatives as part of an overlay strategy, investors can also expand the universe of currencies in their portfolios, over and above their asset-related exposures. By expanding the breadth of currencies in an overlay, expected value-add improves as the manager has a better chance of finding opportunities.

Implications of an overvalued USD

After surging in 2022, the US dollar is now well above its long-term fair value versus peers. By our estimates, it stands at its most overvalued level against G5 currencies since the 1980s. Against the Canadian dollar, we estimate the USD to be 10% above fair value (see page 8).

We expect US inflation to stick above target as the labour market remains resilient to higher interest rates, forcing the Fed to keep rates "higher for longer" and putting a soft floor under the USD. But over the course of next year and beyond, the US dollar's extreme overvaluation should drag it lower — it can only fight gravity for so long. Plus, in the unexpected event that the Fed pivots, the US dollar could quickly revert lower towards its fair value.

Given the large US share of global equity markets — US stocks make up about 60% of the MSCI All-Country World Index — the USD's current overvaluation has major implications for non-US investors. Investors may wish to consider dynamically hedging FX exposures to take advantage of an expected long-term weakening of the US dollar from its current over-valued level.

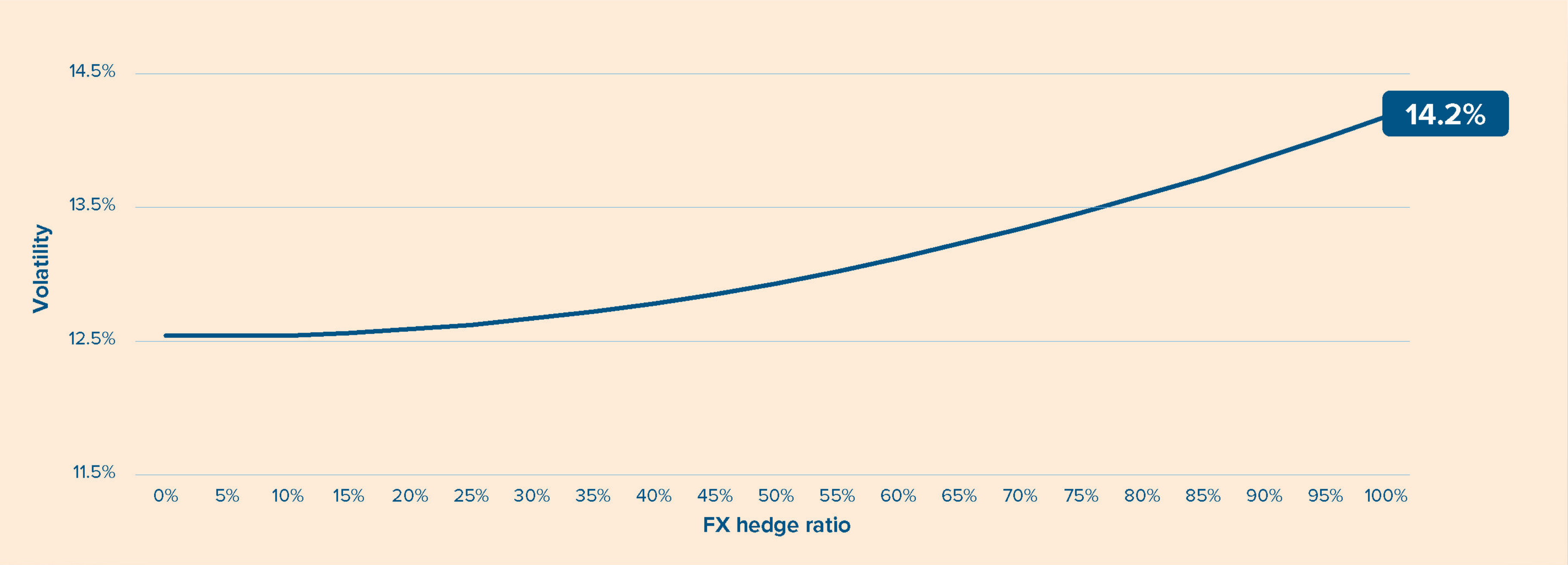


Optimal strategic hedging

Currency hedging decisions can impact the volatility of a diversified global equity portfolio for Canadian-resident investors. Historically, a low hedge ratio for the US dollar (below 30%) has tended to minimize total risk of a foreign equity portfolio.

Volatility of MSCI World Index

based on different FX hedge ratios for Canadian investors

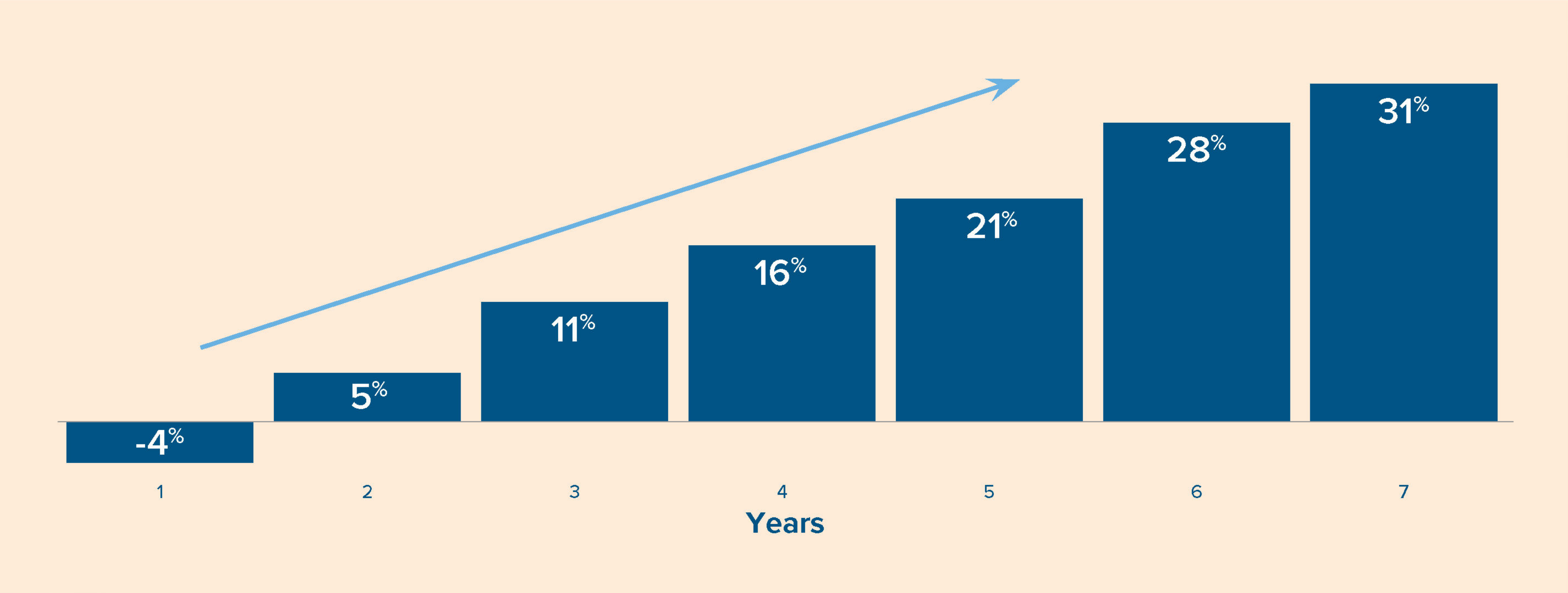


Calculations by the Mackenzie Multi-Asset Strategies Team. Based on Arruda, Bergeron and Kritzman, “Optimal Currency Hedging: Horizon Matters”, Journal of Alternative Investments, 2021.

The optimal FX hedge ratio may depend on the horizon. Over a two-to-seven-year measurement horizon, an unhedged portfolio had greater risk than a hedged one. For investors with short investment horizons, an unhedged portfolio is optimal, while partial hedging can be risk-reducing for investors with longer investment horizons.

Percentage of portfolio risk due to currencies over various measurement horizons

Canadian investor in unhedged international equity



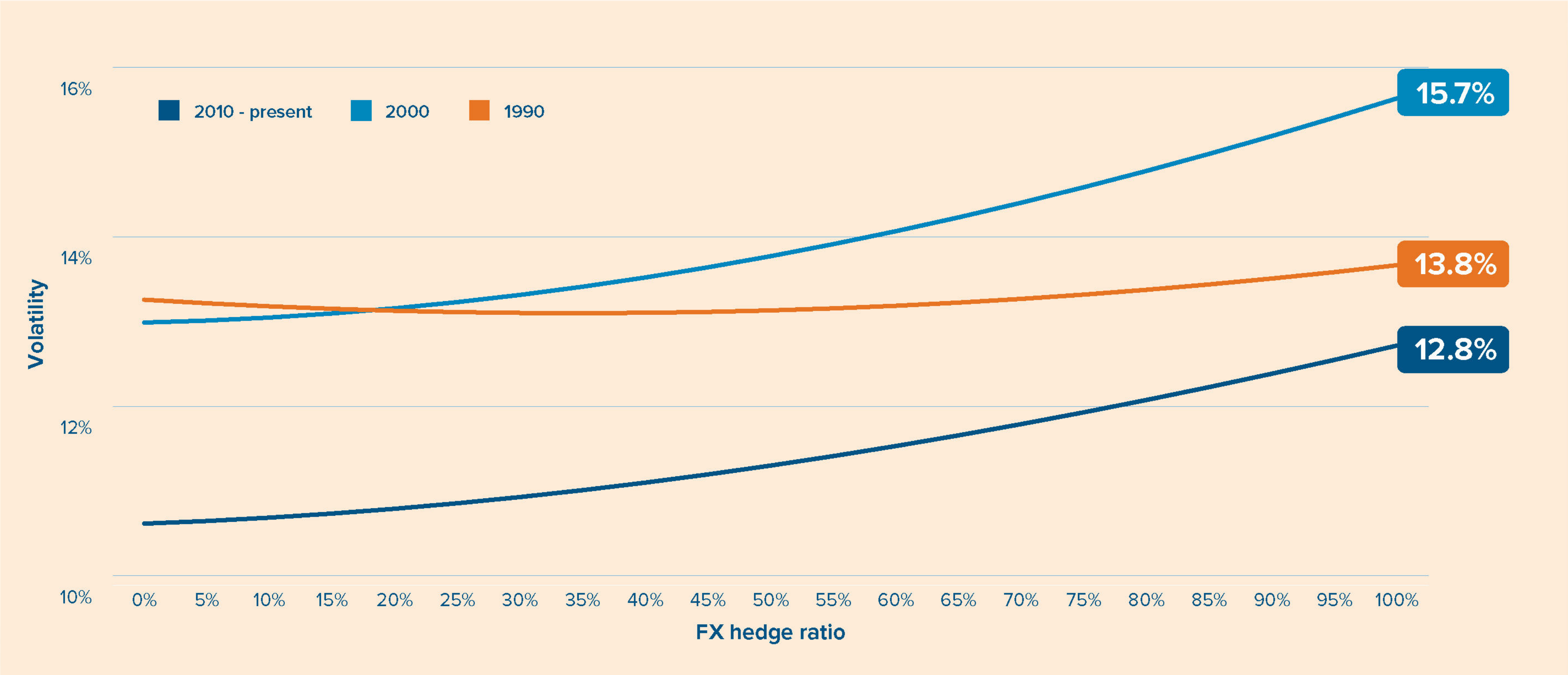
Calculations by the Mackenzie Multi-Asset Strategies Team. Based on Arruda, Bergeron and Kritzman, “Optimal Currency Hedging: Horizon Matters”, Journal of Alternative Investments, 2021.

Dynamic FX hedging

Optimal currency hedging decisions have evolved through time, as correlation regimes change. Currency correlations have differed by decade as macroeconomic conditions in certain countries evolved over time.

Volatility of MSCI World Index for Canadian investors

based on different FX hedge ratios, by decade

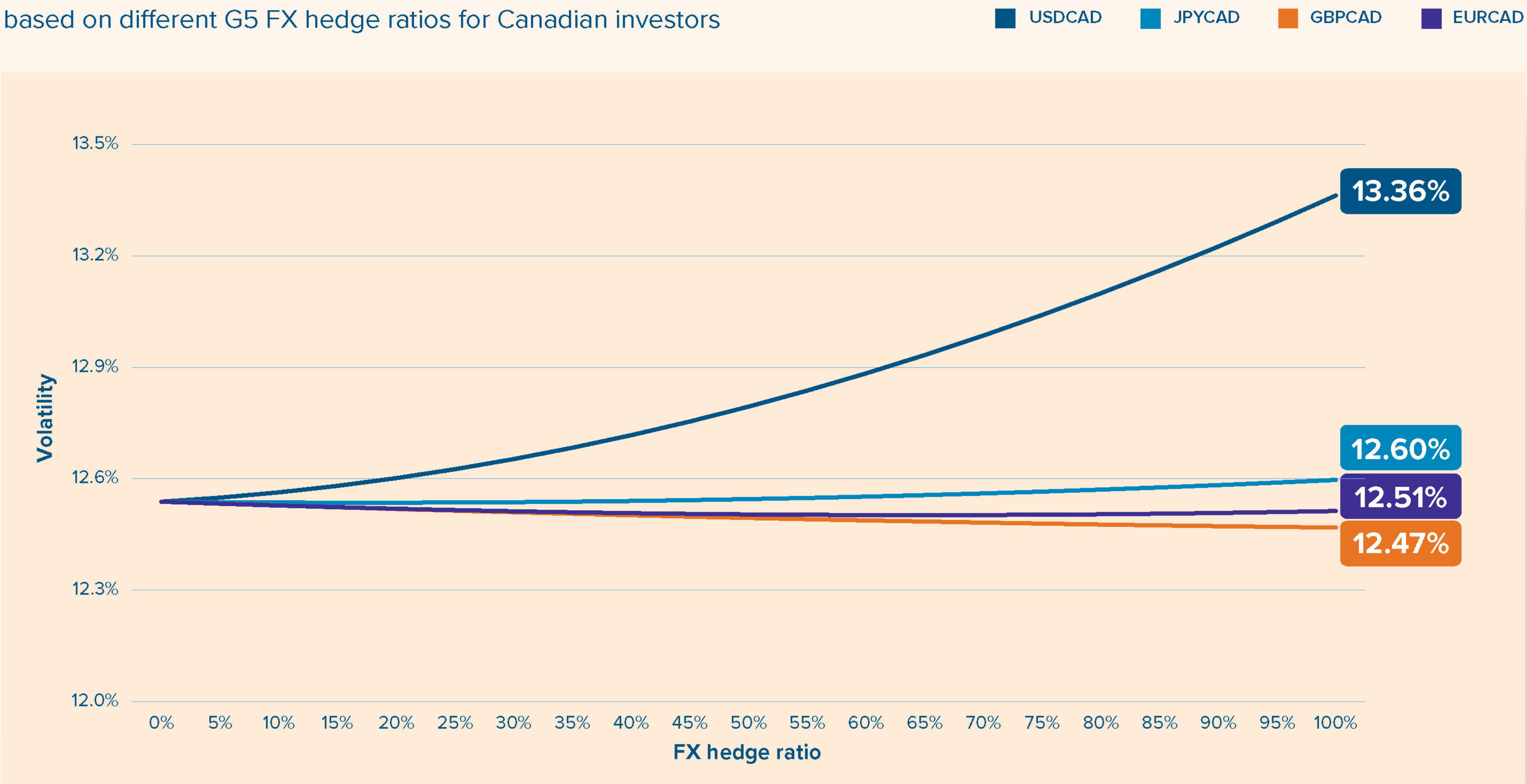


Calculations by the Mackenzie Multi-Asset Strategies Team.

The optimal currency hedge ratio also depends on the specific currency being hedged, as each currency brings different risk characteristics to a portfolio. Unhedged US dollar exposure generally reduces risk for a stock portfolio, while exposure to a more cyclical currency, like the British pound, will tend to increase overall portfolio volatility.

Volatility of MSCI World Index

based on different G5 FX hedge ratios for Canadian investors

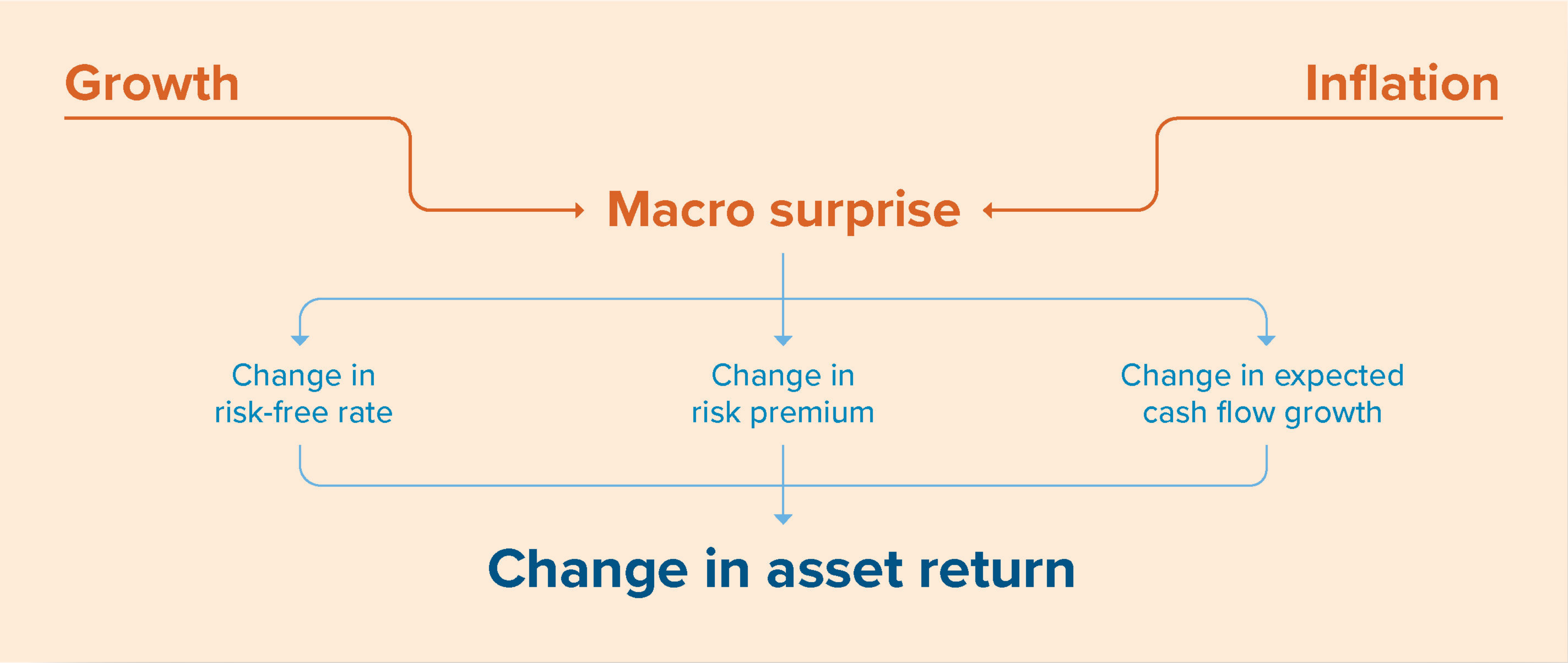


Calculations by the Mackenzie Multi-Asset Strategies Team.

Macroeconomic factors

Long-term expected returns are mainly explained by risk-free rates, unconditional (long-term) risk premiums and starting valuations. But a significant portion of realized expected returns are driven by macroeconomic shocks. For example, China's demand slowdown, commodity oversupply and USD strength were the primary explanation for the disappointing realized EM equity returns in the 2010s.

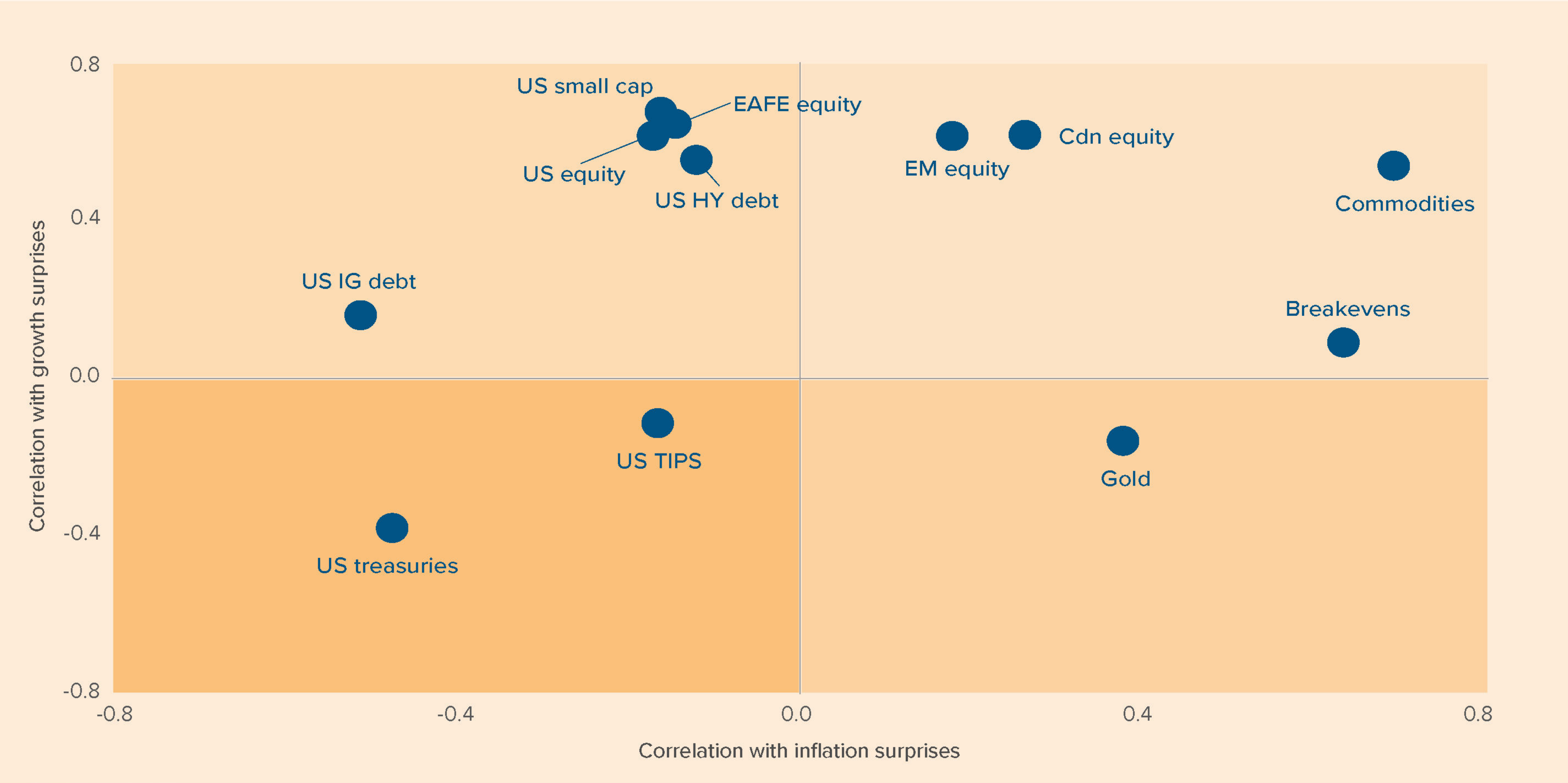
While changes in these macro trends are always difficult to forecast with certainty, we can still estimate the conditional response of asset returns given a macroeconomic shock. This framework for conditional returns, or scenario analysis, can be useful to investors seeking to understand the magnitude of macro risk exposures in their portfolios; help size an active view about macro factors;[1] or inform asset allocation for investors with future liabilities linked to macro factors (such as inflation-adjusted pension payouts).



1 See Alain Bergeron, Mark Kritzman and Gleb Sivitsky. "Asset Allocation and Factor Investing: An Integrated Approach", The Journal of Portfolio Management, Vol. 44, Issue 4, Quantitative Special Issue 2018.

To capture causality, our model uses macro "surprises" — shocks to consensus forecasts of macro variables — rather than current readings of the variables.[2] This framework reflects the intuition that while macro views contribute modestly to long-term unconditional expected returns, macro surprises can and do drive a large portion of realized returns over a cycle.

Correlation with macro factors



2 We use an average of two methods: errors in forecasts from the Survey of Consumer forecasters (as in Thapar et al. (2021), "When Stock-Bond Diversification Fails") and changes in the one-year ahead growth and inflation forecasts from Consensus Economics.

Macro scenarios and returns

Asset class sensitivities or betas to inflation and growth shocks allow investors to estimate the exposure of their portfolio to different economic scenarios. For example, a one standard deviation positive growth shock would cause the average pension portfolio (see p. 14) to gain 5.8%, while a positive inflation shock would cause it to lose 3.3%.

	Shock to growth expectations (sds)	Shock to inflation expectations (sds)	US treasuries	US TIPS	US IG debt	US HY debt	US equity	US small cap	Cdn equity	EAFE equity	EM equity	Gold	Com-modities
Positive growth	+1	no shock	-1.0%	-0.1%	0.5%	4.5%	9.0%	12.1%	9.0%	8.2%	11.2%	-3.2%	9.8%
Positive inflation	no shock	+1	-4.8%	-2.4%	-6.1%	-2.7%	-3.4%	-3.6%	3.2%	-2.5%	2.5%	9.2%	18.4%
Demand-led growth	+1	+1	-5.8%	-2.4%	-5.6%	1.8%	5.5%	8.5%	12.2%	5.7%	13.7%	6.0%	28.2%
Stagflation	-1	+1	-3.8%	-2.3%	-6.6%	-7.1%	-12.4%	-15.6%	-5.8%	-10.6%	-8.8%	12.4%	8.6%
Disinflationary growth	+1	-1	3.8%	2.3%	6.6%	7.1%	12.4%	15.6%	5.8%	10.6%	8.8%	-12.4%	-8.6%
Recession	-1	-1	5.8%	2.4%	5.6%	-1.8%	-5.5%	-8.5%	-12.2%	-5.7%	-13.7%	-6.0%	-28.2%

Our methodology employs the historical beta of asset returns with macro surprises, which we interpret as exogenous shocks to returns. We use an average of two methods: errors in forecasts from the Survey of Consumer forecasters (as in Thapar et al. (2021), "When Stock-Bond Diversification Fails") and changes in the one-year ahead growth and inflation forecasts from Consensus Economics.

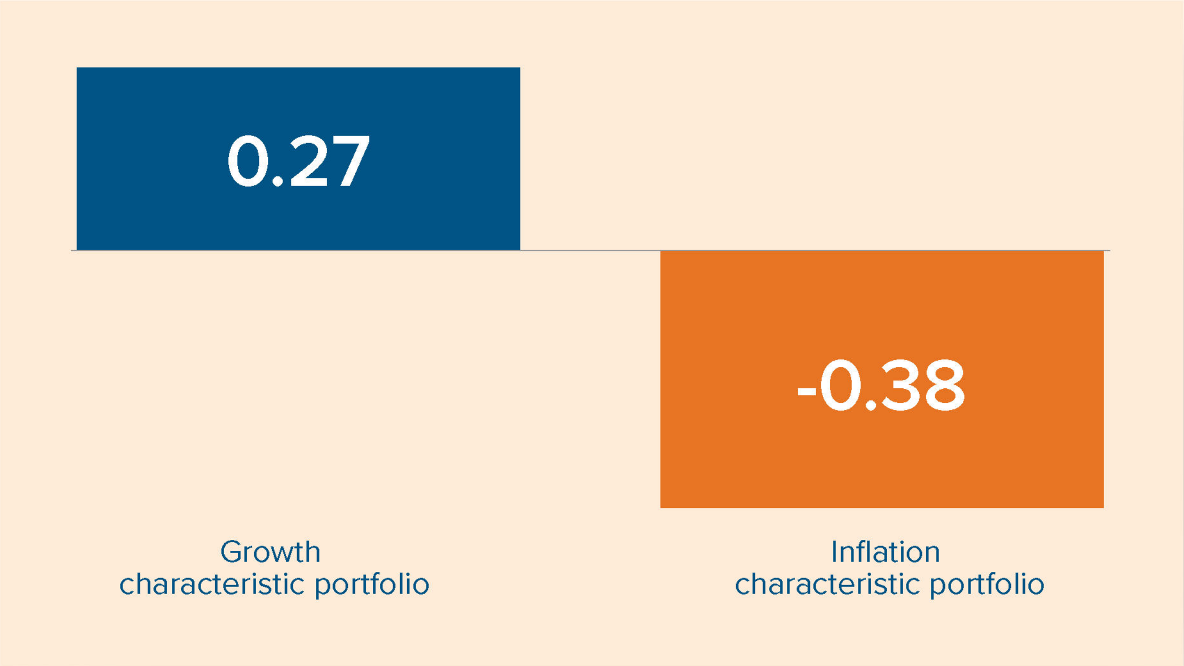
Portfolios of macro factors

In addition to monitoring a portfolio's macro exposures and preventing unwanted risk concentrations, macro factor betas allow investors to explicitly implement macro views in their portfolios. Suppose an investor thinks economic growth will be higher than the market expects. By including a growth factor in the covariance risk matrix, they can size the growth exposure based on their conviction and risk budget. The same framework can be employed to hedge an inflation-sensitive liability, such as pension benefits.

We can also construct a long-short "characteristic" portfolio to represent a pure unit exposure to a macro factor. For example, the returns on the inflation characteristic portfolio on a given day represent shocks to the market's inflation expectations.

The returns of the growth and inflation characteristic portfolios give a hint as to the compensation investors should expect for taking on macro risks. Consumption-based asset pricing theory suggests that assets whose returns exhibit higher correlations with consumption shocks should have higher expected returns. Given consumer utility is positively correlated to growth and negatively to inflation, we would expect a growth characteristic portfolio to have a positive risk-adjusted return and an inflation characteristic portfolio to have a negative risk-adjusted return — that is, investors must "pay for inflation protection". Historical returns support the theory.

Historical Sharpe ratio (1960-today)

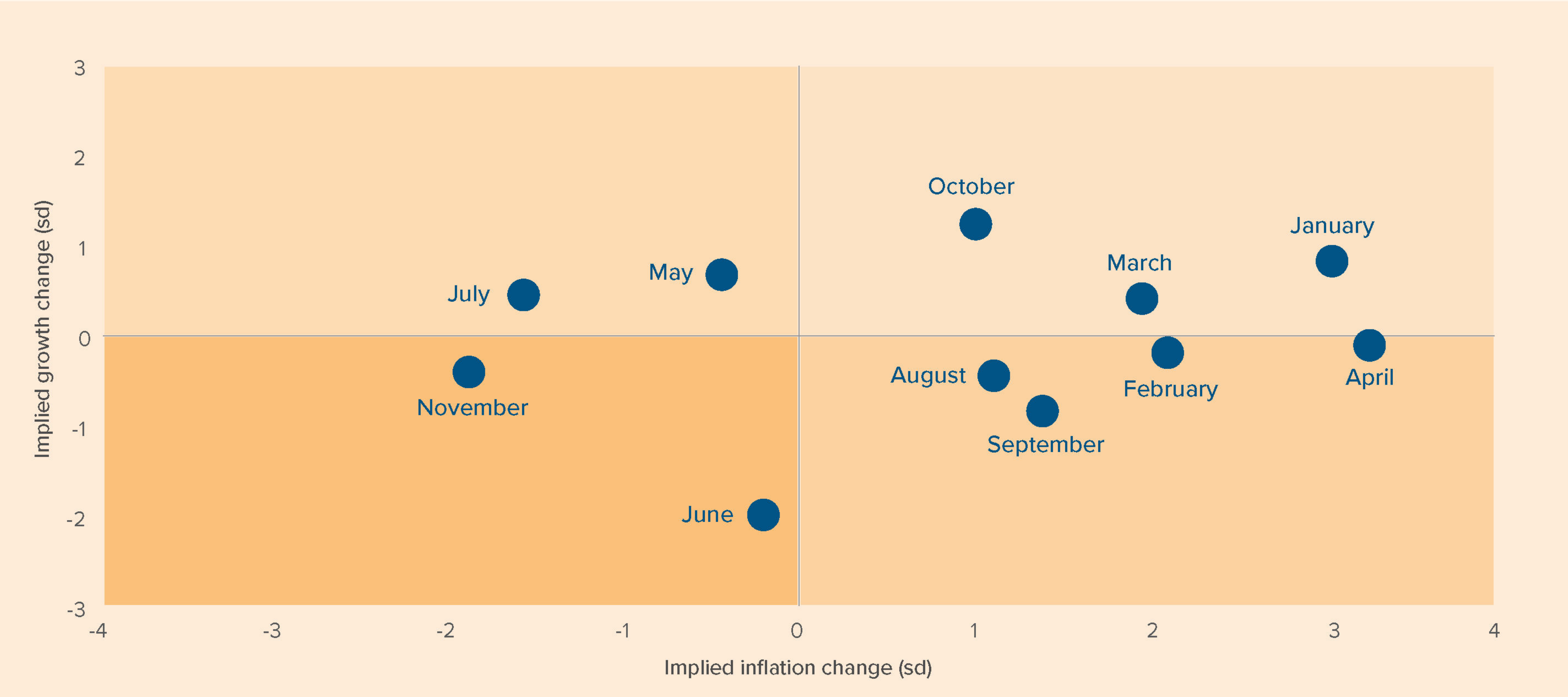


Calculations by the Mackenzie Multi-Asset Strategies Team. The unit characteristic portfolios for growth and inflation are constructed by optimizing the exposure to the macro factor: $h_f = \frac{\Sigma^{-1} f}{f^T \Sigma^{-1} f}$ where Σ is the asset covariance matrix and $f = [f_1, \dots, f_N]$ is a vector representing each asset's exposure to a given macro factor.

Data for the chart via Bloomberg as of November 30, 2022.

The characteristic portfolios for growth and inflation can also act as higher frequency proxies for economic surprises. Economic indicators move slowly and changes in investors' expectations of growth and inflation cannot be observed directly. The returns on the growth and inflation characteristic portfolios can be seen as real-time proxies for shocks to expectations, providing useful information about market expectations as implied by current asset prices.

In 2022, markets were driven more by inflation shocks than growth shocks



Calculations by the Mackenzie Multi-Asset Strategies Team. The unit characteristic portfolios for growth and inflation are constructed by optimizing the exposure to the macro factor: $h_f = \frac{\Sigma^{-1} f}{f^T \Sigma^{-1} f}$ where Σ is the asset covariance matrix and $f = [f_1, \ldots, f_N]$ is a vector representing each asset's exposure to a given macro factor.

Data for the chart via Bloomberg as of November 30, 2022.

Team background

Mackenzie's Multi-Asset Strategies (MAS) Team is co-led by Todd Mattina, Senior Vice President and our in-house Chief Economist, and Nelson Arruda, Senior Vice President and Portfolio Manager. The team has deep expertise across a broad range of strategies including:

- Multi-asset portfolios
- A suite of dynamic currency hedging approaches based on valuation, sentiment and macro conditions developed and maintained in-house
- Liquid alternative strategies that include global macro, commodities, currencies, CTA, and market neutral equity factor portfolios
- Long only, multi-factor equity portfolios (smart beta)

Members of the team engage with institutional investors across Canada on strategic and tactical asset allocation, currency management and engage in academic partnerships to produce thought leadership. The group's pedigree fits naturally with the thinking of institutional investors, and that perspective is reflected in the consideration of risk and return in everything we do.

"The MAS Team manages a broad range of portfolios, including multi-asset funds, dynamic FX hedging strategies, alternatives and multi-factor equity funds.

- Nelson Arruda & Todd Mattina

Todd Mattina
PhD
Co-Lead, Multi-Asset Strategies Team

Nelson Arruda
MFin., MSc., CFA
Co-Lead, Multi-Asset Strategies Team

ORANGE BOOK

2023 LONG-TERM CAPITAL MARKETS OUTLOOK

MACKENZIE

G. S. Maddala and C. R. Rao, eds., *Handbook of Statistics, Vol. 14*

Errors-in-Variables Problems in Financial Models

G. S. Maddala and M. Nimalendran

1. Introduction

The errors-in-variables (EIV) problems in finance arise from using incorrectly measured variables or proxy variables in regression models. Errors in measuring the dependent variables are incorporated in the disturbance term and they cause no problems. However, when an independent variable is measured with error, this error appears in both the regressor variable and in the error term of the new regression model. This results in contemporaneous correlation between the regressor and the error term, and leads to a biased OLS (Ordinary Least Squares) estimator (even asymptotically) and inconsistent standard errors. The biases introduced by measurement errors can be significant and can lead to incorrect inferences. Further, when there are more than one regressor variable in the model the direction of the bias is unpredictable. The effect of measurement errors on OLS estimators is discussed extensively in several econometrics texts including Maddala (1992), and Greene (1993). A comprehensive discussion of errors-in-variables model is in Fuller (1987) and a discussion in the context of econometric models is in Griliches (1985), and Chamberlain and Goldberger (1990).

The errors in the regressor variable could be due to several causes. We can classify them into the following two groups: (1) measurement errors, and (2) use of proxy variables for unobservable theoretical concepts, constructs or latent variables. Measurement errors could be introduced by using estimated values in the regression model. Examples of this are the use of estimated betas as regressors in cross-sectional tests of the CAPM (Capital Asset Pricing Model), and two-pass tests of the APT (Arbitrage Pricing Theory) where estimated rather than actual factor loadings are used in the second pass tests. The second major source of errors arises from the use of proxy variables for unobservable or latent variables. An example of this in finance would be the testing of signaling models where the econometrician observes only a noisy signal of the underlying attribute that is being signaled. In this article we examine several alternative models and techniques employed in financial models to mitigate the errors-in-variables problems. Some areas in finance where errors-in-variables problems are encountered are described below:

I. *Testing asset pricing models*: There are several potential problems in these tests; these include measurement errors associated with the use of estimates for risk measures and the problem associated with the unobservability of the true market portfolio.

II. *Performance measurements*: Measuring the performance of managed portfolios (mutual funds, pension funds etc.) is an important exercise that provides information about the ability of managers to provide superior returns. However, any method used to measure performance must specify a benchmark, and an incorrect specification of the benchmark would introduce errors in the performance measures.

III. *Market response to corporate announcements*: Several articles analyze the response of the market to unexpected earnings, unexpected dividends, unexpected splits and other announcements. To obtain the unexpected component of the variable one needs to specify a model for the expected component. An incorrect specification of the expectation model or estimation errors can result in the unexpected component being measured with error.

IV. *Testing of signaling models*: In signaling models it is argued that managers with private information can employ indicators such as dividends, earnings, splits, capital structure etc. to signal their private information to the market. In testing these models one has to realize that the indicators are noisy measures of the underlying attribute that is signaled (investment opportunities, future cash flows etc.).

A researcher can employ several approaches to correct for the errors-in-variables problem, and to obtain consistent estimates and standard errors. We examine these approaches under the following eight classifications: (1) Grouping Methods, (2) Direct and Reverse Regressions, (3) Alternatives to Two Pass Methods, (4) MIMIC Models, and (5) Artificial Neural Networks (ANN) models. We also discuss other models where the errors-in-variables problems are relevant. These are examined under the categories: (6) Signal Extraction Models, (7) Qualitative Limited Dependent Variable Models, and (8) Factor Analysis with Measurement Errors.

2. Grouping methods

Grouping methods have been commonly used in finance as a solution to the errors-in-variables problem. See, for instance, Black, Jensen and Scholes (1972), Fama and MacBeth (1973) and Fama and French (1992) for a recent illustration. We will refer to these papers as BJS, FM and FF respectively in subsequent discussion. The basic approach involves a two-pass technique. In the first pass, time series data on each individual security are used to estimate betas for each security. In the second pass a cross-section regression (CSR) for the average returns on the securities is estimated using the betas obtained from the first pass as regressors. This introduces the errors-in-variables problem. Since grouping methods can be viewed as instrumental variable (IV) methods, grouping is used to solve this errors-in-variables problem. There are frequent references to Wald's classic paper in this literature but the simple grouping method used by Wald is not the one used in these papers.

Wald's method consists of ranking the observations, forming two groups and then passing a line between the means of the two groups. Later articles suggested that the efficiency of the estimator could be improved by dividing the data into three groups, discarding the observations in the middle group, and passing the line between the means of the upper and lower groups. Wald's procedure amounts to using rank as an instrumental variable, but since rank depends on the measurement error, this cannot produce a consistent estimator (a point noted by Wald himself). Pakes (1982) argues that contrary to the statements often made in several textbooks (including the text by Maddala, 1977, which has been corrected in *Introduction to Econometrics*, Second. Ed. 1992) the grouping estimator is not consistent. This problem has also been pointed out in the finance literature in a recent paper by Lys and Sabino (1992) although there is no reference in this paper to the work of Pakes (1982).

The grouping method used in FM and FF is not the simple grouping method used by Wald. The procedure is to estimate the betas with, say, monthly observations on the first 5 years and then rank the securities based on these estimated betas to form 20 groups (portfolios). Then the estimation sample (omitting the first 5 years of data) is used to estimate a cross-section regression of asset returns on the betas for the different groups.

2.1. *Cross-sectional tests*

In the cross-sectional tests of the CAPM, the average return on a cross-sectional sample of securities over some time period is regressed against each securities beta (β) with respect to a market portfolio. In the first stage, $\hat{\beta}_i$ is estimated from a time series regression of the return on a market index R_{Mt} on the individual stock returns R_{it}.

$$R_{it} = \alpha_i + \beta_i R_{Mt} + v_{it} \ . \tag{1}$$

In the second stage, a cross-sectional regression model of the average return on the individual security $\bar{R}_i$, is regressed on the estimate of beta.

$$\bar{R}_i = \gamma_0 + \gamma_1 \hat{\beta}_i + \epsilon_i \ . \tag{2}$$

Finally, the estimated coefficient $\hat{\gamma}_0$ is compared to the risk-free rate (R_f) in the period under examination and $\hat{\gamma}_1$ is compared to an estimate of the risk premium on the market ($\bar{R}_M - R_f$) estimated from the same estimation period. The first direct test based on cross-sectional regression was by Douglas (1969). In this test Douglas estimated a cross-sectional model of the average return on a large number of common stocks on the stock's own variance and on their covariance with a market index. The tests were inconsistent with the CAPM because the coefficient on the variance term was significant while the coefficient on the covariance term was not significant.

A detailed analysis of the econometric problems that arise from a cross-sectional test was first given by Miller and Scholes (1972). They concluded that measurement error in $\hat{\beta}_i$ was a significant source of bias that contributed toward the findings by Douglas. Fama and MacBeth (1973) use a portfolio approach to reduce the errors-in-variables problem. In particular, they estimate the following cross-sectional-time-series model.

$$R_{pt} = \gamma_{0t} + \gamma_{1t}\beta_{p,t-1} + \gamma_{2t}\bar{\beta}^2_{p,t-1} + \gamma_{3t}\bar{\sigma}_{p,t-1}(\epsilon) + \eta_{pt} \ , \tag{3}$$

where, β_p is the average of the betas for the individual stocks in a portfolio, $\bar{\beta}^2_p$ is the average of the squared betas and $\bar{\sigma}_p(\epsilon)$ is the average residual variance from a market model given by equation (1).

If $\hat{\beta}_i$ is estimated with an unbiased measurement error v_i then the regression estimate of γ for the model described by equation (2) is given by

$$p\lim \quad \hat{\gamma}_1 = \frac{\gamma_1}{1 + \dfrac{\mathrm{Var}\ (v_i)}{\mathrm{Var}\ (\beta_i)}} \tag{4}$$

where, $\mathrm{Var}(v_i)$ is the variance of the measurement errors, and $\mathrm{Var}(\beta_i)$ is the cross-sectional sample variance of the true risk measures β_i. Thus, even for large samples, as long as β_i's are measured with errors the estimated coefficient $\hat{\gamma}_1$ will be biased toward zero and $\hat{\gamma}_0$ will be biased away from its true value. The idea behind the grouping or portfolio technique is to minimize the $\mathrm{var}(v_i)$ through the portfolio diversification effect, and at the same time one would like to maximize the $\mathrm{Var}(\beta_i)$ by forming portfolios by ranking on β_i's.

2.2. *Time series and multivariate tests*

Black, Jensen and Scholes (1972) employ a time-series procedure to test the CAPM that avoids the errors-in-variables problem. They estimate the following model:

$$(R_{pt} - R_{Ft}) = \alpha_p + \beta_p(R_{Mt} - R_{Ft}) + \epsilon_{pt} \ , \tag{5}$$

where, R_{pt} is the return on a portfolio of stocks ranked by their betas estimated from a prior period, R_{Ft} is the risk free rate, and R_{Mt} is the return for the market portfolio. In this specification, the test is based on the hypothesis that $\alpha_p = 0$ if CAPM is valid.

Gibbons (1982) employs a multivariate regression framework in which the asset pricing models are cast as nonlinear parameter restrictions. The approach avoids the errors-in-variables problems introduced by the two pass cross-sectional tests. Gibbons uses the method to test the Black's (1972) version of the CAPM which specifies the following linear relationship between expected return on the security and risk.

$$E(R_{it}) = \gamma + \beta_i[E(R_{mt}) - \gamma] \ , \tag{6}$$

where, $E(R_{it})$ is the expected return on security i for period t, $E(R_{mt})$ is the expected return on the market portfolio for period t, γ is the expected return on a zero beta portfolio, and $\beta_i = \text{cov}(R_{it}, R_{mt})/\text{var}(R_{mt})$. In addition, if asset returns are stationary with a multivariate normal distribution, then they can be described by the "market model"

$$R_{it} = \alpha_i + \beta_i R_{mt} + \eta_{it}, \quad i = 1, \ldots, N \ , \quad t = 1, \ldots T \ . \tag{7}$$

In terms of equation (7), Black's model given by equation (6) implies the restrictions

$$\alpha_i = \gamma(1 - \beta_i) \quad \forall \quad i = 1, \ldots, N \ . \tag{8}$$

Thus, Black's version of the CAPM places nonlinear restrictions on a system of N regression equations. The errors-in-variables problems with the two-pass procedure are avoided by estimating γ and β's simultaneously. Gibbons employs a likelihood ratio statistic to test the restrictions implied by the CAPM.

One important point to note in the cross-sectional tests is that grouping to take care of errors in variable is not necessary. The problem here is not the one in the usual EIV models where the variance of the measurement error is not known. Note that the betas are estimated but their variance is known. This knowledge is used in Litzenberger and Ramaswamy (1979) (referred to later as L-R) to get bias corrected estimates. In the statistical literature this method is known as consistent adjusted least squares (CAL) method and has been discussed by Schneeweiss (1976), Fuller (1980) and Kapteyn and Wansbeek (1984), although the conditions under which the error variances are estimated are different in the statistical literature and the financial literature. The L-R method involves subtracting an appropriate expression from the cross-product matrix of the estimated beta vector to neutralize the impact of the measurement error. The modified estimator is consistent as the number of securities tends to infinity. However, in practice, this adjustment does not always yield a cross-product matrix that is positive definite. In fact, Shanken and Weinstein (1990) observe this in their work and argue that more work is needed on the properties of L-R method. Banz (1981) also mentions "serious problems in applying the Litzenberger-Ramaswamy estimator" in his analysis of the firm size effect.

Besides the L-R method, another promising alternative to the traditional grouping procedure for correcting the EIV bias, is the maximum likelihood method. Shanken (1992) discusses the relationship between the L-R method and the ML method.

In addition to the bias correction problem there is the problem of correcting the standard errors of the estimated coefficients. Shanken (1992) derives the correction factors for the standard errors in the presence of errors-in-variables.

2.3. Grouping in the presence of multiple proxies

The above discussion refers only to simple regression models with one regressor (estimated beta). However, there are models where several regressors are measured with error. Here, grouping by only one variable amount to using only one instrumental variable, and therefore cannot produce consistent estimates. An example of multiple proxies is the paper by Chen, Roll and Ross (1986) which uses the Fama-MacBeth procedure. We will refer to this paper as CRR. They consider five variables describing the economic conditions (monthly growth in industrial production, change in expected inflation, unexpected inflation, term structure, and risk premium measured as the difference between the return on low grade (Baa) bonds and long-term government bonds.) They use a two-pass procedure. In the first pass the returns on a sample of assets are regressed on the five economic state variables over some estimation period (previous five years). On the second pass the beta estimates from the first pass used as independent variables in 12 cross-sectional regressions, one for each of the next 12 months, with asset returns for the month being the dependent variable. Each coefficient in this regression provides an estimate of the risk premium associated with the corresponding state variable. The two-pass procedure is repeated for each year in the sample, yielding time-series estimates of the risk premia associated with the macro variables. The time series means are then tested by a t-test for significant difference from zero.

CRR argue (p. 394) that "to control the errors-in-variables problem that arises from step c of the beta estimates obtained in step b, and to reduce the noise in individual asset returns, the securities were grouped into portfolios." They use size (total market value at the beginning of each test period) as the variable for grouping. CRR further argue that the economic variables were significant in explaining stock returns and in addition these variables are "priced" (as revealed by significant coefficients in the second pass cross-sectional regression). Shanken and Weinstein (1990), however, argue that the CRR results are sensitive to the grouping method used and that the significance of the coefficients in the cross-sectional regression is altered if EIV adjustment is made to the standard errors.

There are two issues that arise in the CRR approach. First, when there are multiple proxies, does grouping by a single variable give consistent estimates? Since grouping by size is equivalent to the use of size as an instrumental variable, what CRR have done is used one instrumental variable (IV). The number of IV's used should be at least equivalent to the number of proxies, in the case of multiple proxies.

The second issue is that of alternatives to the grouping methods. One can use the adjusted least squares as in the L-R method discussed earlier, although there would be the problem of the resulting moment matrix being not positive definite. Shanken and Weinstein (1990) discuss adjusting the standard errors only but (we should be) making adjustments for both the coefficient bias and the standard errors.

3. Alternatives to the two-pass estimation method

In the estimation of the CAPM model, the errors-in-variables problem is created by using the estimated betas in the first stage as explanatory variables in a second stage cross-section regression. Similar problems arise in the two-pass tests of the arbitrage pricing theory (APT) developed by Roll and Ross (1960), Chen (1983), Connor and Korajczyk (1988), Lehmann and Modest (1988) among others.

While Gibbons' (1982) approach avoids the errors-in-variables problem introduced by a two-pass method, the methodology does not address the issue of the unobservability of the "true" market portfolio. As pointed out by Roll (1977), the test of the asset pricing model is essentially a test of whether the proxy used for the "market portfolio" is mean-variance efficient. Gibbons and Ferson (1985) argue that asset pricing models can be tested without observing the "true" market portfolio if the assumption of a constant risk premium is relaxed. This requires a model for conditional expected returns which is used to estimate ratios of betas without observing the market portfolio.

The problems due to the unobservability of the market portfolio and the errors-in-variables problems can be avoided by using one-step methods where the underlying factors are treated as unobservables. We discuss models with unobservables in Section 5, and factor analysis with measurement errors in Section 9.

Geweke and Zhou (1995) provide an alternative procedure for testing the APT without first estimating separately the factors or factor loadings. Their approach is Bayesian. The basic APT assumes that returns on a vector of N assets are related to k underlying factors by a factor model:

$$\begin{aligned} r_{it} &= \alpha_i + \beta_{i1} f_{1t} + \beta_{i2} f_{2t} + \ldots + \beta_{ik} f_{kt} + \epsilon_{it} \ , \\ i &= 1, \ldots, N \ , \quad t = 1, \ldots, T \ , \end{aligned} \tag{9}$$

where, $\alpha_i = \mathrm{E}(r_{it}), \beta_{ik}$ are the factor loadings, and ϵ_{it} are idiosyncratic errors for the i^{th} asset during period t. This model can be written compactly, in vector notation as

$$r_t = \alpha + \beta f_t + \epsilon_t \ , \tag{10}$$

where r_t is an N-vector of returns during period t, α and ϵ_t are $N \times 1$ vectors, f_t is a $k \times 1$ vector and β is a $N \times k$ matrix. The standard assumptions of the factor model are the following:

$$\begin{aligned} &\mathrm{E}(f_t) = 0 \ , \mathrm{E}(f_t f_t') = I \ , \ \mathrm{E}(\epsilon_t | f_t) = 0 \quad \text{and} \\ &\mathrm{E}(\epsilon_t \epsilon_t' | f_t) = \Sigma \ , \quad \text{where} \quad \Sigma = \mathrm{diag}[\sigma_1^2, \ldots, \sigma_N^2] \ . \end{aligned} \tag{11}$$

Also, ϵ_t and f_t are independent and follow multivariate normal distributions.

It has been shown that absence of riskless arbitrage opportunities imply an approximate linear relation between the expected returns and their risk exposure. That is

$$\alpha_i \simeq \lambda_0 + \lambda_1\beta_{1i} + \ldots + \lambda_k\beta_{ki} \qquad i = 1, \ldots, N \ , \tag{12}$$

as $N \to \infty$, where λ_0 is zero-beta rate and λ_k is the risk premium on the k^{th} factor. Shanken (1992) gives alternative approximate pricing relationships under weaker conditions. A much stronger assumption of competitive equilibrium gives the equilibrium version of the APT where the condition (12) is an equality. Existing studies based on the classical methods test only the equillibrium version. Geweke and Zhou (1995) argue that their approach measures the closeness of (12) directly by obtaining the posterior distribution of Q defined as

$$Q = \frac{1}{N}\sum_{i=1}^{N}(\alpha_i - \lambda_0 - \lambda_1\beta_{1i} \ldots - \lambda_k\beta_{ki})^2 \ . \tag{13}$$

For the equilibrium version of APT, $Q \equiv 0$. Geweke and Zhou argue that inference about Q in the classical framework is extremely complicated. They use the Bayesian approach to derive the posterior distribution of Q based on priors for α, β, λ and Σ. Since the Bayesian approach involves the integration of nuisance parameters from the joint posterior distribution and since analytical integration is not possible in this case, they outline a numerical integration procedure based on Gibbs sampling.

The most flexible two-pass approach is the one developed by Connor and Krajezyk (1986, 1988) which is a cross-section approach that can be applied to a large number of assets to extract the factors. By contrast the approach of Geweke and Zhou is a time-series approach and therefore has a restriction on the number of assets that can be considered ($N \leq T - k$). However, the former approach ignores the EIV problem but the latter does not.

Geweke and Zhou illustrate their methodology by using monthly portfolios returns grouped by industry and market capitalization. An important finding is that there is little improvement in reducing the pricing errors by including more factors beyond the first one. (See also the conclusions in Section 9 which argue in favor of fewer factors.)

4. Direct and reverse regression methods

In his 1921 paper in *Metroeconomica*, Gini stated that the slope of the coefficient of the error ridden variable lies between the probability limit of the OLS coefficient and the probability limit of the "reverse" regression estimate of the same coefficient. This result, which has also been derived in Frisch (1934), does not carry over to the multiple regression case in general. This generalization, due to Koopmans (1937), is discussed, with a new proof in Bekker et al. (1985). Apart from Koopmans' proof, later proofs have been given by Kalman (1982) and Klepper and Leamer (1984). It has also been extended to equation systems by Leamer (1987).

All these results require that the measurement errors be uncorrelated with the equation errors. This assumption is not valid in many applications. Erickson

(1993) derives the implications of placing upper and lower bounds on this correlation in a multiple regression model with exactly one mis-measured regressor. Some other extensions of the bounds literature is that by Krasker and Pratt (1986), who use a prior lower bound on the correlation between the proxy and the true regressor, and Bekker et al. (1987) who use as their prior input an upper bound on the covariance matrix of the errors. Iwata (1992) considers a different problem — the case where instrumental variables are correlated with errors. In this case, the instrumental variable method does not give consistent estimates but Iwata shows that tighter bounds can be found if one has prior information restricting the extent of the correlation between the instrumental variables and the regression equation errors.

In the financial literature the effect of correlated errors has been discussed in Booth and Smith (1985). They consider the case where the errors and the systematic parts of both y and x are correlated (all other error correlations are assumed to be zero). They also give arguments as to why allowing for these correlations is important. This analysis has been applied by Rahman, Fabozzi and Lee (1991) to judge performance measurement of mutual fund shares, which depends on the intercept term in the capital asset pricing model. They derive upper and lower bounds for the constant term using direct and reverse regressions. These results on performance measurement are based on the CAPM. There is, however, discussion in the financial literature of performance measurement based on the APT (arbitrage pricing theory) which is a multiple-index/factor model. See Connor and Korajczyk (1986, 1994). In this case, the bounds on performance measurement are difficult to derive. The results by Klepper and Leamer (1984) can be used but they will be based on the restrictive assumption that the errors and systematic parts are uncorrelated (an assumption relaxed in the paper by Booth and Smith). The relaxation of this assumption is important, as argued in Booth and Smith.

5. Latent variables / structural equation models with measurement errors and MIMIC models

5.1. Multiple indicator models

Many models in finance are formulated in terms of theoretical or hypothetical concepts or latent variables which are not directly observable or measurable. However, often several indicators or proxies are available for these unobserved variables. The indicator or proxy variables can be considered as measuring the unobservable variable with measurement errors. Therefore, the use of these indicator variables directly as a regressor variable in a regression model would lead to errors-in-variables problems. However, if a single unobservable (or latent) variable occurs in different equations as an explanatory variable (multiple indicators of a latent variable), then one can get (under some identifiability conditions) consistent estimates of the coefficients of the unobserved variable. These models are discussed in Zellner (1970), Goldberger (1972), Griliches (1974), Joreskog and Goldberger (1975), and popularized by the LISREL program of Joreskog and Sorbom (1989, 1993).[1] Although many problems in finance fall in this category, there are not many applications of these models in finance. Notable exceptions in corporate finance are the models estimated by Titman and Wessels (1990), Maddala, and Nimalendran (1995), and Desai, Nimalendran and Venkataraman (1995).

Titman and Wessels (TW) investigate the determinants of corporate capital structure in terms of unobserved attributes for which they have indicators or proxies which are measured with error. The model consists of two parts: a measurement model, and a structural model which are jointly estimated. In the measurement model, the errors in the proxy variables (e.g. accounting and market data) used for the unobservable attributes are explicitly modeled as follows:

$$X = \Lambda Z + \delta \ . \tag{14}$$

where, $X_{q\times 1}$ is a vector of proxy variables, Z_{mx1} is vector of unobservable attributes and $\Lambda_{q\times m}$ is a matrix of coefficients, and $\delta_{q\times 1}$ is a vector of errors. In the above measurement model, the observed proxy variables are expressed as a linear combination of one or more attributes and a random measurement error. The structural model consists of the relationship between different measures of capital structure (short term debt/equity, long term debt/equity etc.), $Y_{p\times 1}$, and the unobservable attributes Z. The model is specified as follows where ϵ is a vector of errors:

$$Y = \Gamma Z + \epsilon \ . \tag{15}$$

Equations (14) and (15) are estimated jointly using the maximum likelihood technique (estimation techniques are described later in this section). TW estimate the model for 15 proxy variables, 8 attributes and 3 different capital structure variables. In order to identify the model additional restrictions are placed. In particular, it is assumed that the errors are uncorrelated, and 105 of the elements of the coefficient matrix are constrained to be zero. The principal advantage of the above model over traditional regression models is that it explicitly models the errors in the proxy variables. Further, if the model is identified then it can be estimated by full information maximum likelihood (FIML) which gives consistent and asymptotically efficient estimates under certain regularity conditions.

Maddala and Nimalendran [MN] (1995) employ an unobserved components panel data model to estimate the effects of unexpected earnings on change in price, change in bid-ask spreads and change in trading volume. Traditionally, the unexpected earnings (actual-analysts forecast), ΔE, is employed as a regressor in a regression model to explain the changes in spreads (ΔS) or changes in volume

[1] These models have also been discussed extensively under the titles: linear structural models with measurement errors, analysis of covariance structures, path analysis, causal models and content variable models. Bentler and Bonett (1980) and Bollen (1989) provide excellent introductions to the subject.

(ΔV).[2] However, the unexpected earnings are error-ridden proxies for the true unexpected earnings. Therefore, the estimates and the standard errors suffer from all the problems associated with error in variables. MN employ an unobserved components model to obtain consistent estimates of the coefficients on the unobserved variable and the consistent standard errors. In the 3-equation model they consider, it is assumed that the absolute value of the change in price $|\Delta P|$, the change in spread ΔS, and the change in volume ΔV are three indicator variables of the unobserved absolute value of the unexpected true earnings $|\Delta E^*|$. The specification of the model is,

$$\begin{aligned} |\Delta P| &= \alpha_0 + \alpha_1 |\Delta E^*| + \epsilon_1 \\ \Delta S &= \beta_0 + \beta_1 |\Delta E^*| + \epsilon_2 \\ \Delta V &= \gamma_0 + \gamma_1 |\Delta E^*| + \epsilon_3 \ , \end{aligned} \tag{16}$$

where it is assumed that the errors, $\epsilon_i, I = 1,2,3$, are uncorrelated and they are also uncorrelated with the unobserved variable $|\Delta E^*|$. Then the covariance matrix of the observed variables implied by the model is given by

$$\sum = \begin{pmatrix} \alpha_1^2\sigma_e^2 + \sigma_1^2 & \alpha_1\beta_1\sigma_e^2 + \sigma_{12} & \alpha_1\gamma_1\sigma_e^2 + \sigma_{13} \\ - & \beta_1^2\sigma_e^2 + \sigma_2^2 & \beta_1\gamma_1\sigma_e^2 + \sigma_{23} \\ - & - & \gamma_1^2\sigma_e^2 + \sigma_3^2 \end{pmatrix} , \tag{17}$$

where, $\sigma_{ij} = \mathrm{cov}(\epsilon_i, \epsilon_j), i, j = 1,2,3$ and $\sigma_\epsilon^2 = \mathrm{Var}(\Delta E^*)$.

Since the sample estimates of the variance-covariance matrix are consistent estimates of the population parameters, one can estimate the parameters $\alpha_1, \beta_1, \gamma_1, \sigma_1^2, \sigma_3^2$, and σ_e^2, by setting the sample estimates equal to the population variance-covariance elements. However, there are seven unknown parameters and only six pieces of sample information. Therefore the system is under identified and only β_1/α_1 and γ_1/α_1 that are estimable. The parameters α_1, β_1, and γ_1 are not separately estimable. Among the variances $\sigma_1^2, \sigma_2^2, \sigma_3^2$ are estimable and so is $\alpha_1^2\sigma_e^2$. Let the variance-covariance matrix based on sample data be given by

$$S = \mathrm{Var}\begin{pmatrix} |\Delta P| \\ \Delta S \\ \Delta V \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ - & s_{22} & s_{23} \\ - & - & s_{33} \end{pmatrix} . \tag{18}$$

Then consistent estimates for the parameters are given by:

[2] Morse and Ushman (1983) examined a sample of OTC (Over the Counter) firms and found no evidence of change in the spread around earnings announcements. Skinner (1991) using a sample of NASDAQ firms found only a weak evidence of an increase in spread prior to an earnings announcements. Skinner used change in price around the earnings announcement as a proxy for the forecast errors.

$$\frac{\hat{\beta}_1}{\hat{\alpha}_1} = \frac{s_{23}}{s_{13}}, \quad \frac{\hat{\gamma}_1}{\hat{\alpha}_1} = \frac{s_{23}}{s_{12}}, \quad \hat{\alpha}_1^2\hat{\sigma}_e^2 = \frac{s_{12}}{\hat{\beta}_1/\hat{\alpha}_1}, \hat{\sigma}_1^2 = s_{11} - \hat{\alpha}_1^2\hat{\sigma}_e^2,$$
$$\hat{\sigma}_2^2 = s_{22} - \hat{\beta}_1^2\hat{\alpha}_1^2\hat{\sigma}_e^2, \quad \text{and} \quad \hat{\sigma}_3^2 = s_{33} - \hat{\gamma}_1^2\hat{\alpha}_1^2\hat{\sigma}_e^2 \tag{19}$$

It should also be noted that the model described by equations (16) can be written as:

$$\Delta S = \beta_0^* + \frac{\beta_1}{\alpha_1}|\Delta P| + \epsilon_2^*,$$
$$\Delta V = \gamma_0^* + \frac{\gamma_1}{\alpha_1}|\Delta P| + \epsilon_3^*, \quad \text{where,} \tag{20}$$
$$\beta_0^* = \beta_0 - \frac{\beta_1}{\alpha_1}\alpha_0 \quad \text{and} \quad \epsilon_2^* = \epsilon_2 - \frac{\beta_1}{\alpha_1}\epsilon_1 \ .$$

with γ_0^* and ϵ_3^* defined similarly. From equations (19) and (20), it is easy to see that $\hat{\beta}_1/\hat{\alpha}_1$ is the IV (instrumental variable) using ΔV as an instrumental variable, and $\hat{\gamma}_1/\hat{\alpha}_1$ is the IV estimator from using ΔS as an instrumental variable.

The above model shows that it is not necessary to observe the unobservable variable to estimate the parameters of the model. The sample moments contain sufficient information to identify the structural parameters. Also, since the above model is exactly identified, the method-of-moment estimators are also maximum likelihood estimates under normality assumption, with all its desirable properties. The above model gives estimates of the effects of unexpected earnings on the other variables that are free of the errors-in-variables bias involved in studies that use $|\Delta E|$ or $|\Delta P|$ as a proxy for $|\Delta E^*|$. MN find that errors-in-variables can result in *substantial* biases in OLS estimates leading to *incorrect inferences.*

Maddala and Nimalendran (1995) also estimate a 4-equation model in which the absolute value of the unexpected earnings ($|\Delta E|$) is used as an additional proxy. When there are more than 3 indicator variables, the model is over identified (assuming that the errors are mutually uncorrelated and they are uncorrelated with the latent variable). That is there are more unique sample pieces of information than unknown parameters. If there are N indicators then there are $N(N+1)/2$ sample moments (variances and covariances) but there are only $2N$ unknown parameters. The additional information allows one to estimate additional parameters such as some of the covariances between error terms. More importantly, MN use the panel data structure (quarterly earnings for a cross-section of firms) to obtain within group and between group estimates that provide information about the short term and long term effects of earnings surprises on microstructure variables.

5.2. *Testing signaling models*

The study of the relationship between signals and markets' response to them is an important area of financial research. In these models it is argued that managers with private information employ indicators such as dividends, earnings, splits, capital structure etc. to convey their private information to the market. In testing these models one has to realize that the indicators are only "error ridden" proxies for the "true" underlying attribute being signaled. Therefore, the latent variable/ structural equation models would be more suitable compared to the traditional regression models.

Israel, Ofer and Siegel (1990) discuss several studies that use changes in equity value as a measure of the information content of an event (earnings announcement, dividend announcement, etc.) and use this as an explanatory variable in other equations. See, for instance, Ofer and Siegel (1987). All these studies test the null hypothesis that there is no information content about earnings embodied in a given announcement, by testing for a zero coefficient on the change in equity value ΔP. Israel, et.al. assume that ΔP is a noisy measure of the true information content ΔP^*, and they investigate the power of standard tests of hypotheses by simulation for given values of the slope coefficient, and the ratio of the error variance to var(ΔP).

The information in dividend announcements above that in earnings data, and whether such announcements lead to subsequent changes in earnings estimates, have been studied *interalia* in Aharony and Swary (1980) and Ofer and Siegel (1987). Ofer and Siegel use change in equity value surrounding the dividend announcement as a proxy for the information content and use this as an explanatory variable in the dividend change equation. However, a more reasonable model to estimate, that is free of the errors-in-variables bias is to treat information content as an unobserved signal and use change in equity value, unexpected dividends, and change in expected earnings as functions of the unobserved signal. This is illustrated in the paper by Desai, Nimalendran and Venkataraman [DNV] (1995). DNV estimate a latent variable/structural equation model to examine the information conveyed by stock splits which are announced contemporaneously with dividends. They also examine whether dividends and stock splits convey a single piece of information or whether they provide information about more than a single attribute. Their analysis shows that dividends and splits convey information about two attributes, and more importantly the latent variable approach gives unbiased and asymptotically efficient estimators.

Several recent papers in the area of signaling have argued that management may use a combination of signals to reduce the cost of signaling. It is also possible that management can signal in a sequential manner using insider trading and cash dividends (see for example John and Mishra (1990) and the references in it). Many of the signals used by management are changes in dividends, stock splits, stock repurchases, investment and financial policies, insider trading and so on. In testing these models one has to measure the price reaction around the announcement date and also estimate the unexpected component of the signal used (such as unexpected component of dividend change). Generally simple models such as setting the expected dividend equal to past dividend is used. These naive models can lead to substantial errors.

5.3. MIMIC models

If there are multiple indicators and multiple causes, then these models are called MIMIC models (Joreskog and Goldberger (1975)). Note that the multiple indicators of a single or multiple latent variables model is a special case of the MIMIC model. The structural form is

$$\begin{aligned} Y &= \Lambda z^* + \epsilon \\ z^* &= X'\Delta + v \end{aligned} \tag{21}$$

where, $Y_{m\times 1}$ represents the vector of indicator variables, z^* is unobservable and is related to several causes given by the vector $X_{k\times 1}$, and $\Delta_{k\times 1}$ is a vector of parameters. A potential application of the above model in financial research involves the effects of trading mechanisms (or information disclosure) on liquidity and cost of trading. One function of a stock market is to provide liquidity. Several theoretical and empirical papers have addressed this issue (see for example Grossman and Miller (1988), Amihud and Mendelson (1986), Christie and Huang (1994)). The effect of market structure on liquidity is generally examined by analyzing the change in spreads (effective or quoted) associated with stocks that move from one market to another (as in Christie and Huang (1994). However, spread is only one of several proxies that measure liquidity (other proxies are volume of trade, market depth, number of trades, time between trades etc.) More important, there could be several causes driving a stock's liquidity that include: an optimum price, trading mechanism, frequency and type of information, type of investors, type of underlying assets or investment opportunities of the firm. Given multiple indicators and multiple causes, a MIMIC model is more suitable to evaluate effects of trading mechanism and market structure on liquidity.

5.4. Limitations with MIMIC / latent variable models

5.4.1. Problem of poor proxies and choice of proxies

There are several limitations of the latent variable or MIMIC models. Since the model formulation amounts to using the proxies as instrumental variables in the equations other than the one in which it occurs, the problem of poor proxies is related to the problem of poor instrumental variables, on which there is now considerable literature. Therefore the problems associated with the use of poor instruments suggests that caution should be exercised in employing too many indicators. For instance, Titman and Wessels (1988) use 15 indicators and impose 105 restrictions on the coefficient matrix. The problems arising from poor instruments are not likely to be revealed when one includes every conceivable indicator variable in the model.

Very often there are several proxy variables available for the same unobserved variable. For instance, Datar (1994) investigates the effect of 'liquidity' on equity returns. He considers two proxies for liquidity: volume of trading, and size (market value). Apart from the shortcoming that his analysis is based on size-based and volume-based grouping (which amounts to using the proxy variables as instrumental variables), he argues for the choice of volume as the preferred proxy for liquidity based on conventional t-statistics. The problem of choosing between different proxy varibales cannot be done within the framework of conventional analysis. A recent paper by Zabel (1994) analyzes this problem within the framework of likelihood ratio tests for non-nested hypotheses. However, instead of formulating the problem as a choice between different proxies, it would be advisable to investigate how best to use all the proxies to analyze the effect of say "liquidity" on stock returns. This can be accomplished by using the MIMIC model (or multiple indicator model) approach.

Standard asymptotic theory leads us to expect that a weak instrument will result in a large standard error, thus informing us that there is not much information in that variable. However, in small samples a weak instrument can produce a small standard error and a large t-statistic which can be spurious. Dufour(1994) argues that confidence intervals based on asymptotic theory have zero probability coverage in the weak instrument case. The question of how to detect weak instruments in the presence of several instruments is an unresolved issue. There are some studies like Hall, Rudenbusch and Wilcox (1994) that discuss this but this study also relies on an asymptotic test. Jeong (1994) suggests alternative criteria based on an exact distribution. Thus the issue of which indicators to use and which to discard in MIMIC models needs further investigation. It might often be the case that there are some strong theoretical reasons in favor of some indicators and these any how need to be included (as done in the study by DNV).

5.4.2. Violation of assumptions

The second important limitation arises from the assumption that the errors are uncorrelated with the systematic component and among themselves. In the multiple indicator models, some of the correlations among the errors or the errors and the systematic parts may be introduced only if the number of indicators is more than three. The third problem arises from possible non-normality of the errors. In this case the estimates are still consistent, but the standard errors and other test statistics are not valid. Browne(1984) suggests a weighted least squares (WLS) approach which is asymptotically efficient, and provides the correct standard errors and test statistic under general distributional assumption. Finally, there is the question of small sample performance for the different tests based on the latent varibale model and FIML.

5.5. Estimation

All the models described in this section can be estimated by FIML. See Aigner and Goldberger (1977), Aigner, Hsiao, Kapteyn, and Wansbeek (1984), and Bollen (1989). The FIML approach provides an estimator that is consistent, asymptotically efficient, scale invariant, and scale free. Further, through the Hessian matrix one can obtain standard errors for the parameter estimates. However, these standard errors are consistent only under the assumption that the observed variables are multivariate normal. If the observed variables have significant excess kurtosis, the asymptotic covariance matrix, standard errors, and the χ^2 statistic (for model evaluation) based on the estimator are incorrect (even though the estimator is still consistent). Under these conditions, the correct standard errors and test statistics can be obtained by using the asympotically distribution free WLS estimators suggested by Browne (1984). The FIML estimates for the model are obtained by maximizing the following likelihood function.

$$L(\theta) = \text{constant} - \left(\frac{N}{2}\right)\left[\log|\Sigma(\theta)| + \text{tr}[S\Sigma^{-1}(\theta)]\right] \ , \tag{22}$$

where S is the sample variance-covariance matrix for the observed variables, and $\Sigma(\theta)$ is the covariance matrix implied by the model. Several statistical packages including LISREL and SAS provide FIML estimates and their standard errors. LISREL also provides the asymtotically distribution free WLS estimates.

6. Artificial neural networks (ANN) as alternatives to MIMIC models

One other limitation of the models considered in the previous section is the assumption of linearity in the relationships. The artificial neural network (ANN) approach is similar in structure to the MIMIC models (apart from differences in terminology) but allows for unspecified forms of non-linearity. In the ANN terminology the input layer corresponds to the causes in the MIMIC models, and the middle or hidden layer corresponds to the unobservables. In principle, the model can consist of several hidden or middle layers but in practice there is only one hidden layer. The ANN models were proposed by cognitive scientists as flexible non-linear models inspired by certain features of the way the human brain processes information. These models have only recently received attention from statisticians and econometricians. Cheng and Titterington (1994) provide a statistical perspective and Kuan and White (1994) provide an econometrics perspective. An introduction to the computational aspects of these models can be found in Hertz et. al. (1991) and the relationship between neural networks and non-linear least squares in Angus (1989).

The ANN is just a kind of black box with very little said about the nature of the non-linear relationships. Because of their simplicity and flexibility and because they have been shown to have some success compared with linear models, they have been used in several financial applications for the purpose of forecasting. See Trippi and Turban (1993), Kuan and White (1994) and Hutchinson, Lo and Poggio (1994). Apart from the linear vs. nonlinear difference, another major difference is that the MIMIC models have a structural interpretation, but the ANN models do not. However, for forecasting purposes detailed specifications of the structure may not be important. There is considerable discussion about identification in the case of ANN, but the whole emphasis is on approximation and forecasting with a black box. Hornik, Stinchcombe and White (1990), for instance, show that single hidden layer multi-layer neural networks can approximate the derivatives of an arbitrary non-linear mapping arbitrarily well as the number of hidden units increases. Most of the papers on ANN appear in the journal *Neural Newtorks*. However, not much work has been done on comparing MIMIC models discussed in the previous section with ANN models (with the exception of Qi, 1995).

7. Signal extraction methods and tests for rationality

The signal extraction problem is that of predicting the true values for the error-ridden variables. In the statistical literature this problem has been investigated by Fuller (1990). In the finance literature the problem has been discussed by Orazem and Falk (1989). The set-up of the two models is, however, different.

This problem can be analyzed within the context of MIMIC models discussed in the previous section. Consider, for instance, the problem analyzed by Maddala and Nimalendran (1995). Suppose we now have a proxy ΔE for ΔE^* which can be described by the equation,

$$\Delta E = \Delta E^* + \epsilon_4 \ , \tag{23}$$

where, ΔE is unanticipated earnings from say the IBES survey. The estimation of the MIMIC model considered in the previous section gives us an estimate of Var (ΔE^*). The signal extraction approach gives us an estimate of ΔE^* as

$$\Delta \hat{E}^* = \gamma(\Delta E) \qquad \text{where} \qquad \gamma = \frac{\text{Var}\,(\Delta E^*)}{\text{Var}\,(\Delta E)} \ . \tag{24}$$

Thus, if we have a noisy measure of ΔE^*, then this, in conjuction with the other equations in which ΔE^* occurs as an explanatory variable, enables us to get estimates of γ and this can be accomplished if we have other variables where ΔE^* occurs as an explanatory variable. This method can also be used to test rationality of earnings forecasts (say those from the IBES survey). For an illustration of this approach see Jeong and Maddala (1991).

8. Qualitative and limited dependent variable models

Qualitative variable models and limited dependent variable models also fall in the category of unobserved variable models. However, in these cases there is partial observability (observed in a range or in a qualitative fashion). The unobserved variable models discussed in the previous section are of a different category. There is, however, a need to combine the two approaches in the analysis of event studies. For instance, in the signaling models, there are different categories of signals: dividends, stock splits, stock repurchases, etc. In connection with these models there are the two questions, of whether or not to signal, and how best to signal. When considering the information content of different announcements, (say dividend change or stock split) it is customary to consider only the firms that have made these signals. But given that signaling is an endogenous event (the firm has decided to signal), there is a selection bias problem in the computation of abnormal returns computed at the time of the announcement (during the period of the announcement window).

There are studies such as McNichols and Dravid (1990) that consider a matched sample and analyze the determinants of dividends and stock splits. However, the computation of abnormal returns does not make any allowance for the endogeneity of the signals. In addition, there are some conceptual problems involved with the "matched sample" method almost universally used in financial research of this kind. The problem here is the following. Suppose we are investigating the determinants of dividends. We have firms that pay dividends and we get a "matched sample" of firms that do not pay dividends. The match is based on some attribute X that is common to both. Usually the variable X is also used as an explanatory variable in a (logit) model to explain the determinants of dividends. If we have a perfect match, then we have the situation that one firm with the value of X has paid a dividend, and another with the same value of X has not. Obviously, X cannot explain the determinants of dividends. The determinants of dividend payments must be some other variables besides the ones that we use to get matched samples.

The LISREL program can deal with ordinal and censored variables besides continuous variables. However, combining MIMIC models with selection bias in the more relevant financial applications, as in the example of McNichols and Dravid (1990) is more complicated if we allow for endogeneity of the signals. It is, however, true that the self-selection model, has as its reduced form a censored regression model. Thus the LISREL program can be used to account for selection bias *in its reduced form*. But the estimation of MIMIC models with selection bias in the structural form needs further work.

9. Factor analysis with measurement errors

In the econometrics testing of the APT (arbitrage pricing theory) many investigators have suggested that the unobserved factors might be equated with observed macro economic variables. See *inter alia* Chen, Roll and Ross (1986); Chan, Chen and Hsieh (1985); and Conway and Reinganum (1988). The papers using observed variables to represent the factors treat these variables as accurate measures of a linear transformation of the underlying factors so that the regression coefficients are estimates of the factor loadings. However, these observed macro-economic variables are only proxies which at best measure the factors subject to errors of measurement.

Cragg and Donald (1992) develop a framework for testing the APT considering the fact that the factors are measured with error. They apply this technique to monthly returns over the period 1971–90 (inclusive) for 60 companies selected at random form the CRSP tape. They consider 18 macroeconomics variables but found that they represent only four or five factors. The method they used, as outlined in Cragg and Donald (1995) is based on the GLS approach to factor analysis, which is an extension of earlier work by Joreskög and Goldberger (1972) and Dahm and Fuller (1986). Cragg and Donald argue that there is no way of estimating the underlying factors in an APT model without measurement error. In particular this holds for macro-economic variables that are possible proxies. However, as argued in the previous sections, an alternative method to handle the measurement error problem is to use the unobserved components model where the macroeconomic variables (used as proxies) are treated as indicators of unobserved factors. The LISREL program can be used to estimate this model. Tests of the APT can be conducted within this framework as well, and it will be free of the errors-in-variables problem. The LISREL program handles both the GLS and ML estimation methods. However, the MIMIC models impose more structure than the Cragg-Donald approach. A comparison of the two approaches – the multiple indicator approach and the approach of factor analysis with measurement errors is a topic for further research.

10. Conclusion

This article surveys several problems in financial models caused by errors-in-variables and use of proxies. In addition, the article also examines alternative models and techniques that can be employed to mitigate the problems due to errors-in-variables. As noted in the different places, several important gaps exist in the financial literature.

First, many models in finance use grouping methods to mitigate error-in-variables problems. This approach can be viewed as the use of instrumental variable (IV) methods. Therefore, it is appropriate to make use of the recent econometrics literature on instrumental variables, which discusses the problem of poor instruments, judging instrument relevance, and choice among several instruments.

Second, since the use of proxy variables for unobservables is also very pervasive, use can be made of the vast econometrics literature on latent and unobservable variables. For instance, MIMIC models are not used as often as they should be. Also, the interrelationships and comparative performance of MIMIC models, ANN models and factor analytic models with measurement errors need to be studied.

References

Aharony, J. and I. Swary (1980). Quarterly dividend and earnings announcements and stockholders' returns: An empirical analysis. *J. Finance* **35**, 1–12.

Aigner, D. J. and A. S. Goldberger eds., 1977. *Latent Variables in Socio-Economic Models.* North Holland, Amsterdam.

Aigner, D. J., C. Hsiao, A. Kapteyn and T. Wansbeek (1984). Latent variable models in econometrics. In: Z. Griliches and M. D. Intrilligator eds., *Handbook of Econometrics* Vol II, North Holland, 1321–1393.

Amihud, A. R. and H. Mendelson (1986). Asset pricing and the bid-ask spread. *J. Financ. Econom.* **17**, 223–249.

Angus, J. E. (1989). On the connection between neural network learning and multivariate non-linear least squares estimation. *Neural Networks* **1**, 42–47.
Banz, R. (1981). The relations between returns and market values of common stocks. *J. Financ. Econom.* **9**, 3–18.
Bekker, P., A. Kapteyn, and T. Wansbeek (1985). Errors in variables in econometrics: New developments and recurrent themes. *Statistica Neerlandica* **39**, 129–141.
Bentler, P. M. and D. G. Bonett (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* **88**, 588–606.
Black, F., M. C. Jensen and M. Scholes (1972). The capital asset pricing model: Some empirical tests. In: M. Jensen ed., *Studies in the Theory of Capital Markets,* Praeger, New York, 79–121.
Bollen, K. A., (1989). Structural equations with latent variables. New York, Wiley.
Booth, J. R. and R. L. Smith (1985). The application of errors-in-variables methodology to capital market research: Evidence on the small-firm effect. *J. Financ. Quant. Anal.* **20**, 501–515.
Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *Brit. J. Math. Statist. Psych.* **37**, 62–83.
Chamberlain, G. and A. S. Goldberger (1990). Latent variables in econometrics. *J. Econom. Perspectives* **4**, 125–152.
Chan, K. C., N. F. Chen and D. A. Hsieh (1985). An exploratory investigation of the firm size effect. *J. Financ. Econom.* **14**, 451–471.
Chen, N. F., R. Roll, S. A. Ross (1986). Economic forces and the stock market. *J. Business* **59**, 383–403.
Cheng, B. and D. M. Titterington (1994). Neural networks: A review from the statistical perspective (discussion). *Statist. Sci.* **9**, 2–54.
Chen, N. (1983). Some empirical tests of the theory of arbitrage pricing. *J. Finance* **38**, 1392–1414.
Christie, W. G. and R. D. Huang (1994). Market structures and liquidity: A transactions data study of exchange listings. *J. Finan. Intermed.* **3**, 300–326.
Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory. *J. Financ. Econom.* **15**, 373–394.
Connor, G. and R. A. Korajczyk (1988). Risk and return in an equilibrium APT: An application of a new methodology. *J. Financ. Econom.* **21**, 255–289.
Connor, G. and R. A. Korajczyk (1994). Arbitrage pricing theory. In: R. Jarrow, V. Maksimovic, and W.T. Ziemba eds., *The Finance Handbook*, North Holland Publishing Co.
Conway, D. A. and M. C. Reinganum (1988). Stable factors in securing returns: Identification using cross-validation. *J. Business Econom. Statist.* **6**, 1–15.
Cragg, J. G. and S. G. Donald (1992). Testing and determining arbitrage pricing structure from regressions on macro variables. *University of British Columbia*, Discussion paper #14.
Cragg, J. G. and S. G. Donald (1995). Factor analysis under more general conditions with reference to heteroskedasticity of unknown form. In: G. S. Maddala, Peter Phillips and T. N. Srinivasan eds., *Advances in Econometrices and Quantative Economics*, Essays in Honor of C. R. Rao (Blackwell).
Datar, V. (1994). Value of liquidity in financial markets. Unpublished Ph.D. dissertation, University of Florida.
Desai, A. S., M. Nimalendran, and S. Venkataraman (1995). Inferring the information conveyed by multiple signals using latent variables/structural equation models. Manuscript, *University of Florida*, Department of Finance, Insurance and Real Estate.
Dahm, P. F. and W. A. Fuller (1986). Generalized least squares estimation of the functional multivariate linear errors in variables model. *J. Multivar. Anal.* **19**, 132–141.
Douglas, G. W. (1969). Risk in the equity markets: An empirical appraisal of market efficiency. *Yale Economic Essays* **9**, 3–45.
Dufour, J. M. (1994). Some impossibility theorems in econometrics with applications to instrumental variables, dynamic models and cointegration. Paper presented at the Econometric Society European Meetings, Maastricht.
Erickson, T. (1993). Restricting regression slopes in the errors-in-variables model by bounding the error correlation. *Econometrica* **61**, 959–969.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *J. Finance* **47**, 427–465.
Fama, E. F. and J. MacBeth (1973). Risk, return and equilibrium: Empirical tests. *J. Politic. Econom.* **81**, 607–636.
Frisch, R. (1934). *Statistical Confluence Analysis by Means of Complete Regression Systems.* Oslo, University Institute of Economics.
Fuller, W. A. (1990). Prediction of true values for the measurement error model. In: P. J. Brown and W. A. Fuller eds., *Statistical Analysis of Measurement Error Models and Applications: Contemporary Mathematics* Vol. **12**, 41–58.
Fuller, W. A. (1980). Properties of some estimators for the errors-in-variables model. *Ann. Statist.* **8**, 407–422.
Geweke, J. and G. Zhou (1995). Measuring the pricing error of the arbitrage pricing theory. Federal Reserve Bank of Minneapolis, Research Dept., Staff report #789.
Gibbons, M. R. (1982). Multivariate tests of financial models, a new approach. *J. Financ. Econom.* **10**, 3–27.
Gibbons, M. R. and W. Ferson (1985). Testing asset pricing models with changing expectations and an unobservable market portfolio. *J. Financ. Econom.* **14**, 217–2236.
Goldberger, A. S. (1972). Structural equation methods in the social sciences. *Econometrica*, **40**, 979–1001.
Greene, W. H., (1993). *Econometric Analysis*, 2nd ed., Macmillan, New York.
Griliches, Z. (1974). Errors in variables and other observables. *Econometrica* **42**, 971–998.
Griliches, Z. (1985). Economic data issues. In: Z. Griliches and M. D. Intrilligator eds., *Handbook of Econometrics*, Vol III, North Holland, Amsterdam.
Grossman, S. J. and M. H. Miller (1988). Liquidity and market structure. *J. Finance* **43**, 617–637.
Hall, A. R., G. D. Rudenbusch and D. W. Wilcox (1994). Judging instrument relevance in instrumental variable estimation. Federal Reserve Board, Washington D. C.
Hertz, J., A. Krogh, and R. G. Palmer (1991). *Introduction to the Theory of Neural Computation.* Addison Welsey, Redmont City.
Hornik, K., M. Stinchcombe and H. White (1990). Universal approximation of an unknown mapping and its derivatives. *Neural Networks* **3**, 551–560.
Hutchinson, J. M., A. M. Lo and T. Piggo (1994). A non-parametric approach to pricing and hedging derivative securities via learning networks. *J. Finance* **49**, 851–899.
Israel, R., A. R. Ofer and D. R. Siegel (1990). The use of the changes in equity value as a measure of the information content of announcements of changes in financial policy. *J. Business Econom. Statist.* **8**, 209–216.
Iwata, S. (1992). Instrumental variables estimation in errors-in-variables models when instruments are correlated with errors. *J. Econometrics* **53**, 297–322.
Jeong, J. (1994). On pretesting instrument relevance in instrumental variable estimation. Unpublished paper, Emory University.
Jeong, J. and G. S. Maddala, (1991). Measurement errors and tests for rationality. *J. Business Econom. Statist.* **9**, 431–439.
John, K. and B. Mishra (1990). Information content of insider trading around corporate announcements: The case of capital expenditures. *J. Finance* **45**, 835–855.
Jöreskog, K. G. and A. S. Goldberger (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *J. Amer. Statist. Assoc.* **70**, 631–639.
Jöreskog, K. G. and D. Sorböm (1989). LISREL 7. User's Reference, (First Ed.), SSI Inc. Publication, Chicago.
Jöreskog, K. G. and D. Sorböm (1993). LISREL 8. Structural equation modeling with the Simplis™ command language. SSI Inc. Publication, Chicago.
Kalman, R. E. (1982). System identification from noisy data. In: A. Bednarek and L. Cesari eds., *Dynamical Systems* II, New York Academic Press.
Kapteyn, A. and T. Wansbeek (1984). Errors in variables: Consistent adjusted least squares (CALS) estimation. *Communications in Statistics: Theory and Methods* **13**, 1811–37.
Klepper, S. and E. E. Leamer (1984). Consistent sets of estimates for regression with errors in all variables. *Econometrica* **55**, 163–184.

Koopmans, T. C. (1937). *Linear Regression Analysis of Economic Time Series*. Haarlem, Netherlands Economic Institute, DeErven F. Bohn, NV.
Krasker, W. S. and J. W. Pratt (1986). Bounding the effects of proxy variables on regression coefficients. *Econometrica* **54**, 641–655.
Kuan, C. M. and H. White (1994). Artificial neural networks: An econometric perspective. *Econom. Rev.* **13**, 1–91.
Leamer, E. (1987). Errors in variables in linear systems. *Econometrica* **55**, 893–909.
Lehmann, B. N. and D. M. Modest (1988). The empirical foundations of the arbitrage pricing theory. *J. Financ. Econom.* **21**, 213–254.
Litzenberger, R. H. and K. Ramaswamy (1979). The effect of personal taxes and dividends on capital asset prices. *J. Financ. Econom.* **7**, 163–195.
Lys, T. and J. S. Sabino (1992). Research design issues in grouping-based tests. *J. Financ. Econom.* **32**, 355–387.
Maddala, G. S. (1992). *Introduction to Econometrics*. 2nd ed., Macmillan, New York.
Maddala, G. S. and M. Nimalendran (1995). An unobserved component panel data model to study the effect of earnings surprises on stock prices, volume of trading and bid-ask spreads. *J. Econometrics* **68**, 299–242.
McNichols, M. and A. Dravid (1990). Stock dividends, stock splits, and signaling. *J. Finance* **45**, 857–879.
Miller, M and M. Scholes (1972). Rates of returns in relation to risk: A reexamination of some recent findings. In: M. Jensen ed., *Studies in the Theory of Capital Markets*, Praeger, New York, 47–78.
Morse, D. and N. Ushman (1983). The effect of information announcements on market microstructure. *Account. Rev.* **58**, 274–258.
Ofer, A. R. and D. R. Siegel (1987). Corporate financial policy, information, and market expectations: An Empirical investigation of dividends. *J. Finance* **42**, 889–911.
Orazem, P. and B. Falk (1989). Measuring market responses to error-ridden government announcements. *Quart. Rev. Econom. Business* **29**, 41–55.
Pakes, A. (1982). On the asymptotic bias of the Wald-type estimators of a straight-line when both variables are subject to error. *Internat. Econom. Rev.* **23**, 491–497.
Qi, M. (1995). A comparative study of Neural Network and MIMIC Models in a study of option pricing. Working Paper, Ohio State University.
Rahman, S., F. J. Fabozzi, and C. F. Lee (1991). Errors-in-variables, functional form, and mutual fund returns. *Quart. Rev. Econom. Business.* **31**, 24–35.
Roll, R. W. (1977). A critique of the asset pricing theory's tests-part I: On past and potential testability of the theory. *J. Financ. Econom.* **4**, 129–176.
Roll, R. W. and S. A. Ross (1980). An empirical investigation of the arbitrage pricing theory. *J. Finance* **35**, 1073–1103.
Schneeweiss, H. (1976). Consistent estimation of a regression with errors in the variables. *Metrika* **23**, 101–115.
Shanken, J. (1992). On the estimation of beta-pricing models. *Rev. Financ. Stud.* **5**, 1–33.
Shanken, J. (1992). The current state of the arbitrage pricing theory. *J. Finance* **47**, 1569–74.
Shanken, J. and M. I. Weinstein (1990). Macroeconomic variables and asset pricing: Further results. University of Southern California.
Skinner, D. J. (1991). Stock returns, trading volume, and the bid-ask spreads around earnings announcements; Evidence from the NASDAQ national market system. *The University of Michigan*
Titman, S. and R. Wessels (1988). The determinants of capital structure choice. *J. Finance* **43**, 1–19.
Trippi, R. and E. Turban (1993). *Neural Networks in Finance and Investing*. Chicago, Probus.
White, H. (1989). Some asymptotic results for learning in single hidden-layer feed forward network models. *J. Amer. Statist. Assoc.* **86**, 1003–1013.
Zabel, J. E. (1994). Selection among non-nested sets of regressors: The case of multiple proxy variables. Discussion paper, Tufts University.
Zellner, A. (1970). Estimation of regression relationships containing unobservable independent variables. *Internat. Econom. Rev.* **11**, 441–454.

MarketWatch

Yes, 100% of economists were dead wrong about yields

By Ben Eisen

Published: Oct 22, 2014 8:01 a.m. ET

Back in April every economist in a survey thought yields would rise. Guess what they did next

Getty Images

As it turns out, economists are not soothsayers.

NEW YORK (MarketWatch) — Just about six months ago, a headline flashed across the top of MarketWatch's home page. It read: "100% of economists think yields will rise within six months."

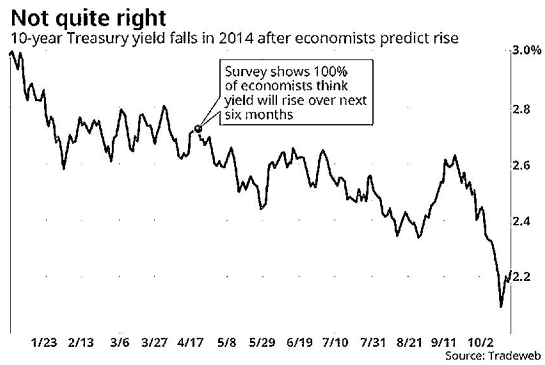
The April 22 report was based on a Bloomberg survey of 67 economists, all of whom expected the 10-year Treasury note 10_YEAR, +0.34% yield — which closed at 2.73% that day — to rise over the following half year.

"How quickly we would get to 4[%] was the discussion at the beginning of the year," said Mohamed El-Erian, chief economic adviser at Allianz SE, on CNBC Tuesday morning.

The market, however, has a funny way of leaning one way, just as the herd is heading in the other direction.

On Tuesday, the 10-year note traded at a yield of 2.21%, almost four-tenths of a percentage point lower than in April. Let's not forget that the yield unexpectedly dipped below 2%, just last week.

That underscores the difficulty of calling the direction of interest rates. It also makes all 67 economists wrong, as this chart of the benchmark yield shows:



Treasury yields tend to rise, and prices drop, as the U.S. economy grows and investors begin to expect the Federal Reserve to normalize monetary policy more quickly.

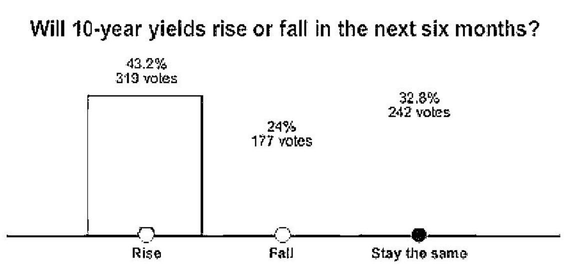
"There's an inherent bias out there that you can only get validation that the economy is improving if rates go up," said George Goncalves, head of interest-rate strategy at Nomura Securities. He was among the strategists saying in the spring that yields would keeping falling.

But the relationship between yields and the economy isn't always linear. Despite steady improvement in the economic numbers, yields have continued to fall. That's in part because of sluggish growth abroad, which has helped push back market views of when the central bank will begin hiking rates.

Goncalves added that falling yields have actually been a boon to the economy this year, keeping financial conditions loose and supporting the housing market. That creates a somewhat paradoxical situation where economic growth and yields are moving in the opposite direction.

The survey of economists' yield projections is generally skewed toward rising rates — only a few times since early 2009 have a majority of respondents to the Bloomberg survey thought rates would fall. But the unanimity of the rising rate forecasts in the spring was a stark reminder of how one-sided market views can become. It also teaches us that economists can be universally wrong.

Then again, the majority of MarketWatch readers weren't exactly expecting rates to fall either, judging by an informal survey taken at the time:



Looking forward, can you guess in which direction the most recent Bloomberg survey of economists shows yields are headed? Yep, the answer is up.

Do you think the 10-year yield will rise or fall in the next six months?

Rise OR Fall

MarketWatch

The Ultimate Poison Pill: Closing the Value Gap

James M. McTaggart, Chairman & Chief Executive Officer

Seldom in the history of U.S. business has a structural change hit with the same force. Ten years ago, large-scale LBOs, raiders, and forced restructuring were virtually unknown. Today, they are commonplace and are rapidly changing the economic landscape. At the source of this structural change is a growing belief that many large diversified companies are not being managed to create the maximum value possible for their shareholders. It is also important to note that the gap between actual and potential market values, the "value gap," is so large for some companies that substantial profits can be made even after premiums of 30- 50% are paid to acquire control. This perception, combined with a flood of institutional money into junk bonds and LBO funds, has produced the takeover entrepreneur, who can now entice or threaten all but the very largest corporations.

Can it be true? Is the value gap of sufficient size to make a large number of diversified companies attractive takeover candidates? In general, the answer is yes, although the number of candidates has been declining recently due to the spread of value-based strategic management. More important, however, are the sources of the gap. There are three management shortcomings that we believe account for most of the gap between actual and potential market values:

1) A tendency to invest far too much capital in unprofitable businesses

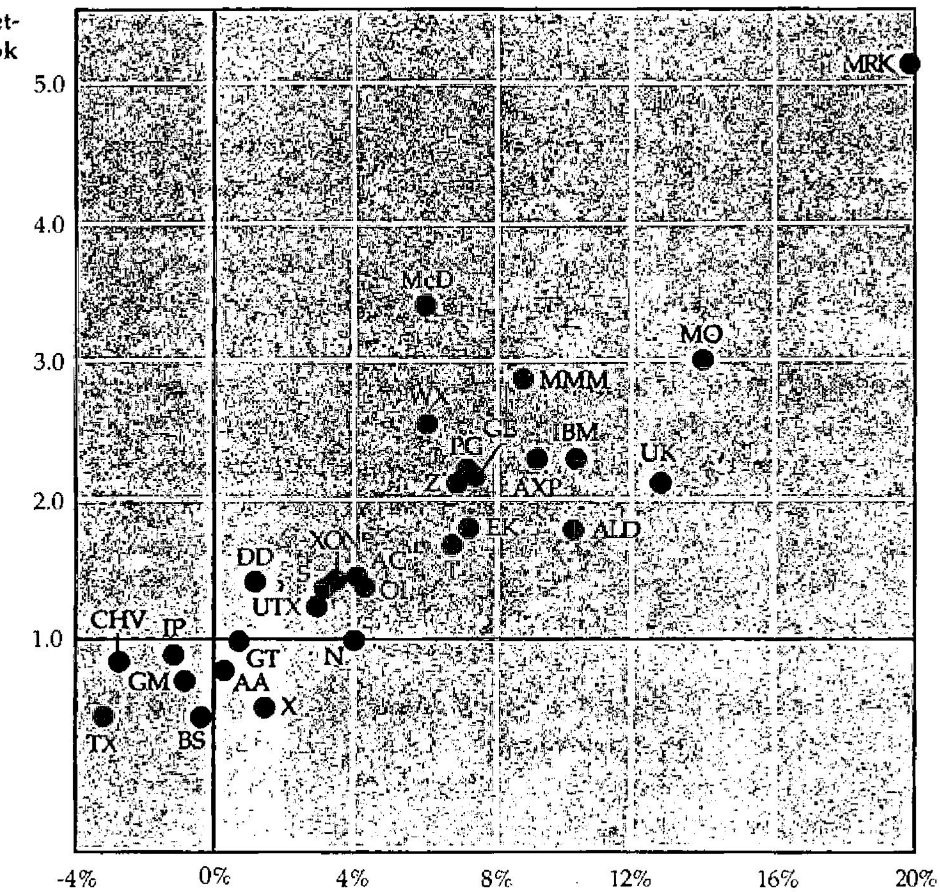
2) Poor balance sheet management, and

3) Tolerance of noneconomic overhead.

The Determinants of Value

In order to describe clearly the three sources of the value gap, it is necessary to first examine the factors that determine the market value of any business or company.

Exhibit 1: Profitability of Dow Jones Industrials - June 1986



ALD	Allied Corp.	IP	Int'l. Paper
AA	Aluminum Co. of Am.	McD	McDonald's Corp.
AC	American Can	MRK	Merck & Co.
AXP	American Express	MMM	Minnesota Mining
T	American Telephone	MO	Philip Morris
BX	Bethlehem Steel	OI	Owens-Illinois
CHV	Chevron	PG	Proctor & Gamble
DD	DuPont	S	Sears, Roebuck
EK	Eastman Kodak	TX	Texaco, Inc.
XON	Exxon Corp.	UK	Union Carbide
GE	General Electric	X	U.S. Steel
GM	General Motors	UTX	United Technologies
GT	Goodyear Tire	WX	Westinghouse
IBM	Int'l. Business Machines	ZZ	Woolworth (F.W.)
N	Inco Limited		

Fundamentally, the value of a company is determined by the cash flow it generates over time for its owners and the minimum acceptable rate of return required by investors to supply equity capital. This "cost of equity capital" is used to discount the expected equity cash flow, converting it to a present value. The cash flow is, in turn, produced by the interaction of a company's return on equity and the annual rate of equity growth. High-ROE companies in low-growth markets, such as Kellogg, are prodigious generators of cash flow, while low-ROE companies in high-growth markets, such as Texas Instruments, barely generate enough cash flow to finance growth.

A company's ROE over time relative to its cost of equity also determines whether it is worth more or less than its book value. If ROE is consistently greater than the cost of equity capital (the investor's minimum acceptable return), the business is economically profitable and its market value will exceed book value. If, however, the business earns an ROE consistently less than its cost of equity, it is economically unprofitable and its market value will be less than book value. These basic principles can be seen at work in **Exhibit I**, which plots the profitability of the Dow Jones Industrials, based on Value Line forecasts of ROE and Marakon estimates of the cost of equity capital.

Growth acts as a magnifier. If ROE remains constant and the growth rate of a profitable business increases, its market-to-book ratio rises. For an unprofitable business, increasing growth actually drives the market-to-book lower (unless growth causes ROE to rise). And in the case where ROE is just equal to the cost of equity, growth has no impact on the market-to-book ratio. The primary reason for the scattering of the observations in **Exhibit I** is differential growth rates.

The profitability of a company is determined primarily by the profitability of its businesses. The profitability of a business is, in turn, determined by economic forces affecting supply and demand in its product markets, its competitive position, and the effectiveness of its strategy. The interaction of constantly changing economic forces and competitive strategies produces a wide variation in both industry and company profitability, as can be seen in **Exhibits II** and **III**. Understanding how industry economics and competitive position determine profitability for a given business is the first step toward developing strategies to increase shareholder returns.

Exhibit II: Profitability of 14 U.S. Industries – Spring 1986

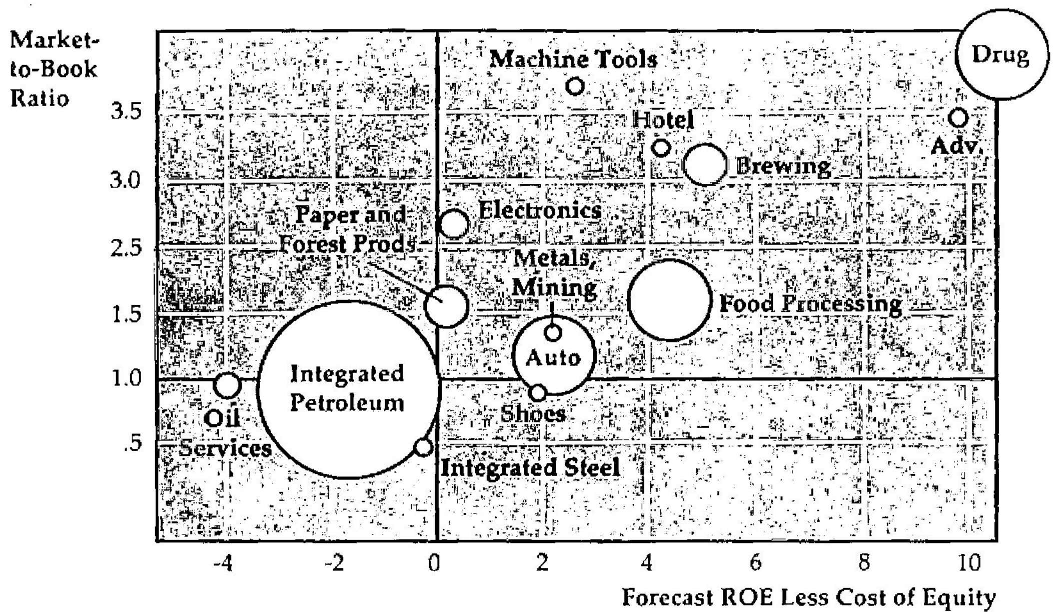
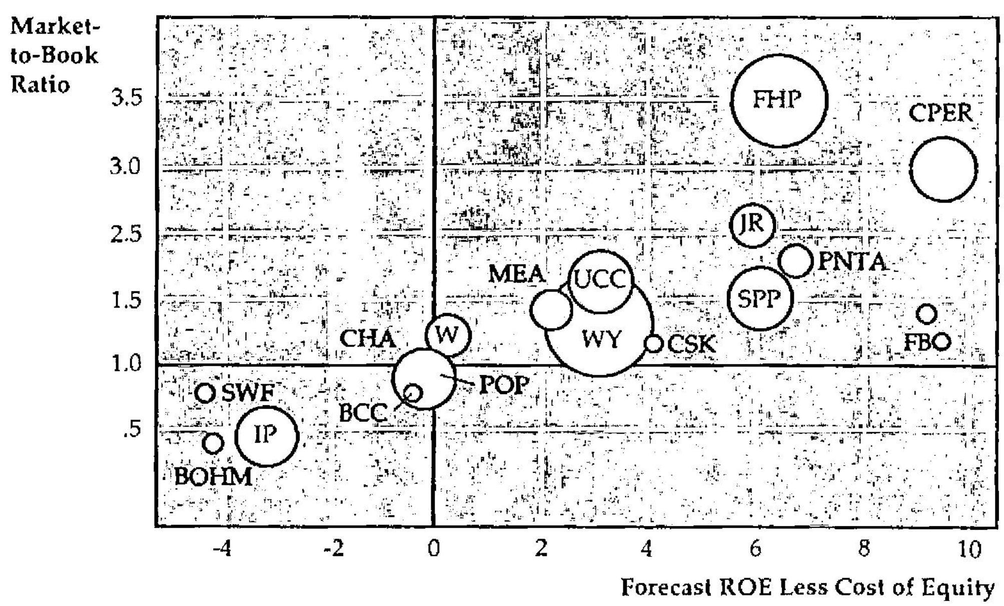


Exhibit III: Profitability of Paper and Forest Products Companies – Spring 1986



BOHM	Bohemia	FHP	Fort Howard Paper	SPP	Scott Paper
BCC	Boise Cascade	IP	International Paper	SWF	Southwest Forest
CHA	Champion International	JR	James River	UCC	Union Camp
CSK	Chesapeake	MEA	Mead	W	Westvaco
CPER	Consolidated Paper	PNTA	Pentair	WY	Weyerhaeuser
FBO	Federal Paper Board	POP	Pope & Talbot		

Sources of the Value Gap

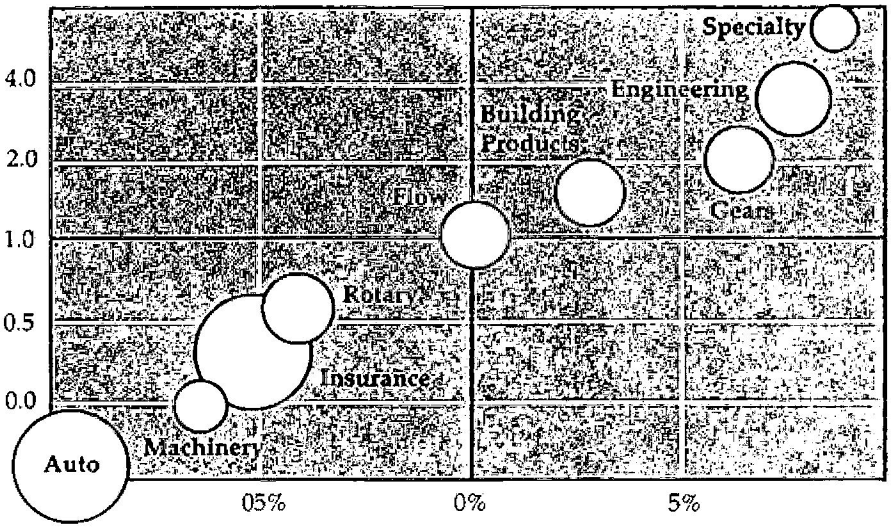
The wide variation in industry and company profitability also occurs within a typical diversified company's portfolio of businesses. Within a company, however, the capital allocation discipline provided by creditors and investors is replaced by management policies and strategies, which can significantly magnify the variation, particularly on the downside. The magnification can occur in either of two ways. The first is when management allows low-return businesses to invest too much capital, a process that can actually produce businesses with negative market values. The second is when management allows or causes high-return businesses to underinvest, which if prolonged usually results in a loss of competitive position and declining returns. In both instances, the business unit market values are significantly lower than they otherwise would be. This tendency to misallocate capital by allowing or causing businesses to pursue inappropriate strategies is the first of the three major sources of the gap between actual and potential market value.

The business portfolio shown in **Exhibit IV**, based on a recent engagement, illustrates the magnitude of the gap that can be produced by pursuing inappropriate business strategies. This company's sales were roughly $750 million, and its common stock was trading at about 80% of book value. Its portfolio contained five profitable and four unprofitable businesses. The operating value of each unprofitable business, based on the prevailing strategies, was less than 50% of its book value. All told, the four operating values summed to $115 million, versus a combined book value exceeding $300 million.

The most unprofitable business, machinery, was actually worth a negative $12 million; that is, the present value of its planned cash flow was negative $12 million. This was produced by an operating strategy whose primary objective was growth. The key element of the plan was a massive capital spending program designed to boost capacity and eliminate a competitive cost disadvantage. And while the program, if successful, would have significantly enhanced the unit's ROI (from 8% to 12%), the long-term positive impact on value was more than offset by the near-term negative cash flow.

Based on a thorough assessment of market economics and profitability relative to competitors, we concluded that by changing strategy at each

Exhibit IV: Profitability of Company Portfolio



of the four businesses to emphasize profitability rather than growth, their combined market values could be increased by at least $150 million within two years. In other words, the current value gap caused by over-investing in four unprofitable businesses was $120 million, or 40% of the company's market value.*

As a general rule, strategy changes at the business unit level emanating from improved capital allocation can enhance market values by anywhere from 20-100% within a few years. While this alone can provide impetus to takeover entrepreneurs, the value gap can, in fact, be further magnified by poor balance sheet management and tolerance of noneconomic overhead.

With respect to balance sheet management, substantial value can often be created by redeploying underperforming assets and reducing the cost of capital used to fund investments. On the asset side, two of the more prominent targets are excess cash and underutilized real estate. The source of value creation in the cash account is the low after-tax return it earns. To the extent that excess cash is held for long periods of time in

*The machinery business was subsequently sold in a leveraged buyout for book value and has since prospered.

taxable securities, it is worth less than its face value. Redeploying excess cash by repurchasing shares, for example, generates a capital gain equal to the present value of the tax savings. Excess pension fund reserves are also a source of funds that can be worth more if returned to shareholders. The source of value creation with corporate real estate is land or buildings that are not being put to their highest and best use. The capital tied up in undeveloped land, vacant office space, underutilized plants, or unprofitable retail outlets nearly always earns a return well below the cost of capital. To the extent that it can be redeployed into profitable businesses or, again, used to buy back stock, a substantial capital will occur.

On the liability side, value can be created for equity holders by increasing financial leverage up to a point. This, of course, is one of the sources of value that LBOs have utilized to recapture purchase price premiums. The source of the value creation is the tax saving due to the deductibility of interest. As a rule of thumb, each dollar of new debt should increase the firm's equity value by 20-25 cents until the firm's financial risk becomes excessive. At this point, the benefits from futher borrowing are offset by the restrictions placed on the firm, which limit its capital availability and increase the probability that the interest expense will not be tax deductible. This point, however, is significantly beyond the current leverage position of most U.S. companies.

The magnitude of the opportunity to increase returns through improved balance sheet management will, of course, depend on the amount of nonproductive assets on the company's books and its capacity to borrow. In the case of Gulf Oil, we estimated that redeployment of over $1 billion of excess cash and full utilization of the company's debt capacity would have produced a 20-25% increase in the market value of Gulf's stock. Focused efforts to reduce underperforming assets and improve liability management can result in increases to shareholder value of up to 50%.

With respect to overhead, our experience suggests that most large companies are overburdened and do not appreciate the magnitude of the overhead drag on equity values. The accumulation of overhead throughout most companies occurs for a variety of reasons. As companies grow, they face the continuing problem of how to decentralize operating responsibility while maintaining some centralized control. In many instances, the result is duplication of support functions at corporate, group, and business unit levels, such as accounting, personnel, and planning. In addition, the overriding objective of most people managing the support functions is to maximize the quality of their services, and their compensation is often closely correlated to the number of people under their stewardship. The result is excess staff and a service "quality-to-cost" ratio that is much lower than it should be.

The impact of noneconomic overhead on value can be staggering. For example, the overhead at Beatrice Corp. was estimated at roughly $150 million annually, or 1.3% of its $12 billion in sales. By contrast, Esmark, at roughly $6 billion in sales, was spending only $25 million on overhead functions, less than 0.5%. If Beatrice could have managed down its overhead to $50 million, the resulting $100 million in pretax earnings would have created roughly $1 billion of shareholder value. This represents nearly 30% of Beatrice's preacquisition market value and 70% of the premium paid to acquire control of the company. This means that if the new owners can manage down Beatrice's overhead to Esmark's level, they will be two thirds of the way to recovering the acquisition premium, with potential divestments, strategy changes, and the impact of leverage and taxes yet to be considered.

Closing the Value Gap

In the current environment, with takeover financing readily available, no company can run for long with a large perceived gap between actual and potential market values. To close the gap, we recommend a five-step process:

First, develop accurate estimates of the operating and divestment values of each business in the portfolio. Few companies have this information, and yet it is the foundation of managing for shareholder value.

Second, incorporate profitability and operating values into both the strategic planning process and incentive compensation. The planning process should stress the relationships among market economics, competitive position, and profitability. Business unit managers cannot be expected to develop value-creating strategies if they don't know how much their units are worth or why they are either profitable or unprofitable. To ensure effective implementation, a significant portion of key executive compensation must be tied directly or indirectly to shareholder value.

Third, don't hoard cash or carry nonproductive assets on the books. At least once a year, a thorough analysis of asset productivity should be conducted.

Fourth, put in place an aggressive financial policy. The level of borrowing should be matched to the ability of business units to bear interest rate risk. Excess cash flow should be dedicated to profitable diversification, dividends, and repurchasing shares.

Fifth, don't tolerate noneconomic overhead. Support functions should be viewed as service businesses and where possible, subjected to both performance measurement and outside competition.

If managed well, a diversified company could be worth more than just the sum of its business unit values, owing to economies of scale and scope in support functions and to the increase in debt capacity produced by diversification. Those companies that can accomplish this feat will not only enrich shareholders but will also put in place the best possible poison pill.

Earnings Forecasts: A Primer

By BEN MCCLURE Updated June 26, 2022

Investopedia

https://www.investopedia.com/articles/stocks/06/earningsforecasts.asp#:~:text=To%20predict%20earnings%2C%20most%20analysts,factors%20that%20influence%20corporate%20growth.

Reviewed by THOMAS BROCK

Fact checked by HANS DANIEL JASPERSON

Anyone who reads the financial press or watches financial news on television will have heard the term "beat the street," which really just means to beat Wall Street earnings forecasts. Wall Street analysts' consensus earnings estimates are used by the market to judge stock performance. Here we offer a brief overview of the consensus earnings and what they mean to investors.

KEY TAKEAWAYS

- Large brokerages hire a slew of analysts to publish reports on various corporations upcoming profit reports, including earnings-per-share and revenue forecasts.
- Consensus earnings estimates refer to the average or median forecasts of a group of analysts as to what a company is expected to earn or lose in a given period of time, typically quarterly and annually.
- While there are some flaws in the system, consensus estimates are perceived as significant in terms of understanding a stock's valuation and are monitored by investors and the financial press.
- Whether a company meets, beats or misses forecasts can have a huge impact on the price of the underlying stock, particularly in the short term.

What Are Consensus Earnings?

Consensus earnings estimates are far from perfect, but they are watched by many investors and play an important role in measuring the appropriate valuation for a stock. Investors measure stock performance on the basis of a company's earnings power. To make a proper assessment, investors seek a sound estimate of this year's and next year's earnings per share (EPS), as well as a strong sense of how much the company will earn even further down the road.

That's why, as part of their services to clients, large brokerage firms—the sell side of Wall Street and other investment communities—employ legions of stock analysts to publish forecast reports on companies' earnings over the coming years.

A consensus forecast number is normally an average or median of all the forecasts from individual analysts tracking a particular stock. So, when you hear that a company is expected to earn $1.50

per share this year, that number could be the average of 30 different forecasts. On the other hand, if it's a smaller company, the estimate could be the average of just one or two forecasts.

A few companies, such as Refinitiv and Zacks Investment Research, compile estimates and compute the average or consensus.

Consensus numbers can also be found at a number of financial websites such as Yahoo! Finance. Estimates are found by looking up individual stocks, for example, Amazon.

Some of these sites also show how estimates get revised upward or downward.

Consensus earnings estimates are not fixed—analysts will typically revise their forecasts as new information comes in, such as company news or regulatory or industry-specific information.

What Time Period Is Covered?

Consensus estimates of quarterly earnings are published for the current quarter, the next quarter and so on for about eight quarters. In some cases, forecasts are available beyond the first few quarters. Forecasts are also compiled for the current and next 12-month periods.

A consensus forecast for the current year is reported once the actual results for the previous year are released. As actual numbers are made available, analysts typically revise their projections within the quarter or year they are forecasting.

Even the most sophisticated investors, including mutual fund and pension fund managers, rely heavily on consensus estimates. Most of them do not have the resources to track thousands of publicly-listed companies in detail, or even to keep tabs on a fraction of them, for that matter.

Why Focus On Earnings?

Many investors rely on earnings performance to make their investment decisions. Stocks are assessed according to their ability to increase earnings as well as to meet or beat analysts' consensus estimates. This influences a company's implicit value (i.e., the personal perceptions and research of investors and analysts), which in turn can affect whether a stock's price rises or drops.

The basic measurement of earnings is earnings per share. This metric is calculated as the company's net earnings—or net income found on its income statement—minus dividends on preferred stock, divided by the number of outstanding shares. For example, if a company (with no preferred stock) produces a net income of $12 million in the third quarter and has eight million shares outstanding, its EPS would be $1.50 ($12 million/8 million).

So, why does the investment community focus on earnings, rather than other metrics such as sales or cash flow? Any finance professor will tell you the only proper way to value a stock is to predict the long-term free cash flows of a company, discount those free cash flows to the present day and divide by the number of shares. But this is much easier said than done, so investors often shortcut the process by using accounting earnings as a "good enough" substitute for free cash flow. Accounting earnings certainly are a much better proxy for free cash flow than sales. Besides, accounting earnings are fairly well defined and public companies' earnings statements must go through rigorous accounting audits before they are released. As a result, the investment community views earnings as a fairly reliable, not to mention convenient, measure. (To read more, see: Getting The Real Earnings.)

What's the Basis of Analysts' Forecasts?

Earnings forecasts are based on analysts' expectations of company growth and profitability. To predict earnings, most analysts build financial models that estimate prospective revenues and costs.

Many analysts will incorporate top-down factors such as economic growth rates, currencies and other macroeconomic factors that influence corporate growth. They use market research reports to get a sense of underlying growth trends. To understand the dynamics of the individual companies they cover, really good analysts will speak to customers, suppliers and competitors. The companies themselves offer earnings guidance that analysts build into the models.

To predict revenues, analysts estimate sales volume growth and estimate the prices companies can charge for the products. On the cost side, analysts look at expected changes in the costs of running the business. Costs include wages, materials used in production, marketing and sales costs, interest on loans, etc.

Analysts' forecasts are critical because they contribute to investors' valuation models. Institutional investors, who can move markets due to the volume of assets they manage, follow analysts at big brokerage houses to varying degrees.

> *Consensus estimates are so consistently tracked by so many stock market players that when a company misses forecasts, it can send a stock tumbling; similarly, a stock that merely meets forecasts might get sent lower, as investors have already priced in the in-line earnings.*

What Are the Implications for Investors?

Consensus estimates are so powerful that even small deviations can send a stock higher or lower. If a company exceeds its consensus estimates, it is usually rewarded with an increase in stock price. If a company falls short of consensus numbers—or sometimes if it only meets expectations—its share price can take a hit.

With so many investors watching consensus numbers, the difference between actual and consensus earnings is perhaps the single most important factor driving share-price performance over the short term. This should come as little surprise to anyone who has owned a stock that "missed the consensus" by a few pennies per share and, as a result, tumbled in value.

For better or for worse, the investment community relies on earnings as its key metric. Stocks are judged not only by their ability to increase earnings quarter over quarter but also by whether they are able to meet or beat a consensus earnings estimate. Like it or not, investors need to keep an eye on consensus numbers to keep tabs on how a stock is likely to perform.

Earnings Forecasts: A Primer

By BEN MCCLURE Updated June 26, 2022

Investopedia

https://www.investopedia.com/articles/stocks/06/earningsforecasts.asp#:~:text=To%20predict%20earnings%2C%20most%20analysts,factors%20that%20influence%20corporate%20growth.

Reviewed by THOMAS BROCK

Fact checked by HANS DANIEL JASPERSON

Anyone who reads the financial press or watches financial news on television will have heard the term "beat the street," which really just means to beat Wall Street earnings forecasts. Wall Street analysts' consensus earnings estimates are used by the market to judge stock performance. Here we offer a brief overview of the consensus earnings and what they mean to investors.

KEY TAKEAWAYS

- Large brokerages hire a slew of analysts to publish reports on various corporations upcoming profit reports, including earnings-per-share and revenue forecasts.
- Consensus earnings estimates refer to the average or median forecasts of a group of analysts as to what a company is expected to earn or lose in a given period of time, typically quarterly and annually.
- While there are some flaws in the system, consensus estimates are perceived as significant in terms of understanding a stock's valuation and are monitored by investors and the financial press.
- Whether a company meets, beats or misses forecasts can have a huge impact on the price of the underlying stock, particularly in the short term.

What Are Consensus Earnings?

Consensus earnings estimates are far from perfect, but they are watched by many investors and play an important role in measuring the appropriate valuation for a stock. Investors measure stock performance on the basis of a company's earnings power. To make a proper assessment, investors seek a sound estimate of this year's and next year's earnings per share (EPS), as well as a strong sense of how much the company will earn even further down the road.

That's why, as part of their services to clients, large brokerage firms—the sell side of Wall Street and other investment communities—employ legions of stock analysts to publish forecast reports on companies' earnings over the coming years.

A consensus forecast number is normally an average or median of all the forecasts from individual analysts tracking a particular stock. So, when you hear that a company is expected to earn $1.50

per share this year, that number could be the average of 30 different forecasts. On the other hand, if it's a smaller company, the estimate could be the average of just one or two forecasts.

A few companies, such as Refinitiv and Zacks Investment Research, compile estimates and compute the average or consensus.

Consensus numbers can also be found at a number of financial websites such as Yahoo! Finance. Estimates are found by looking up individual stocks, for example, Amazon.

Some of these sites also show how estimates get revised upward or downward.

Consensus earnings estimates are not fixed—analysts will typically revise their forecasts as new information comes in, such as company news or regulatory or industry-specific information.

What Time Period Is Covered?

Consensus estimates of quarterly earnings are published for the current quarter, the next quarter and so on for about eight quarters. In some cases, forecasts are available beyond the first few quarters. Forecasts are also compiled for the current and next 12-month periods.

A consensus forecast for the current year is reported once the actual results for the previous year are released. As actual numbers are made available, analysts typically revise their projections within the quarter or year they are forecasting.

Even the most sophisticated investors, including mutual fund and pension fund managers, rely heavily on consensus estimates. Most of them do not have the resources to track thousands of publicly-listed companies in detail, or even to keep tabs on a fraction of them, for that matter.

Why Focus On Earnings?

Many investors rely on earnings performance to make their investment decisions. Stocks are assessed according to their ability to increase earnings as well as to meet or beat analysts' consensus estimates. This influences a company's implicit value (i.e., the personal perceptions and research of investors and analysts), which in turn can affect whether a stock's price rises or drops.

The basic measurement of earnings is earnings per share. This metric is calculated as the company's net earnings—or net income found on its income statement—minus dividends on preferred stock, divided by the number of outstanding shares. For example, if a company (with no preferred stock) produces a net income of $12 million in the third quarter and has eight million shares outstanding, its EPS would be $1.50 ($12 million/8 million).

So, why does the investment community focus on earnings, rather than other metrics such as sales or cash flow? Any finance professor will tell you the only proper way to value a stock is to predict the long-term free cash flows of a company, discount those free cash flows to the present day and divide by the number of shares. But this is much easier said than done, so investors often shortcut the process by using accounting earnings as a "good enough" substitute for free cash flow. Accounting earnings certainly are a much better proxy for free cash flow than sales. Besides, accounting earnings are fairly well defined and public companies' earnings statements must go through rigorous accounting audits before they are released. As a result, the investment community views earnings as a fairly reliable, not to mention convenient, measure. (To read more, see: Getting The Real Earnings.)

What's the Basis of Analysts' Forecasts?

Earnings forecasts are based on analysts' expectations of company growth and profitability. To predict earnings, most analysts build financial models that estimate prospective revenues and costs.

Many analysts will incorporate top-down factors such as economic growth rates, currencies and other macroeconomic factors that influence corporate growth. They use market research reports to get a sense of underlying growth trends. To understand the dynamics of the individual companies they cover, really good analysts will speak to customers, suppliers and competitors. The companies themselves offer earnings guidance that analysts build into the models.

To predict revenues, analysts estimate sales volume growth and estimate the prices companies can charge for the products. On the cost side, analysts look at expected changes in the costs of running the business. Costs include wages, materials used in production, marketing and sales costs, interest on loans, etc.

Analysts' forecasts are critical because they contribute to investors' valuation models. Institutional investors, who can move markets due to the volume of assets they manage, follow analysts at big brokerage houses to varying degrees.

> *Consensus estimates are so consistently tracked by so many stock market players that when a company misses forecasts, it can send a stock tumbling; similarly, a stock that merely meets forecasts might get sent lower, as investors have already priced in the in-line earnings.*

What Are the Implications for Investors?

Consensus estimates are so powerful that even small deviations can send a stock higher or lower. If a company exceeds its consensus estimates, it is usually rewarded with an increase in stock price. If a company falls short of consensus numbers—or sometimes if it only meets expectations—its share price can take a hit.

With so many investors watching consensus numbers, the difference between actual and consensus earnings is perhaps the single most important factor driving share-price performance over the short term. This should come as little surprise to anyone who has owned a stock that "missed the consensus" by a few pennies per share and, as a result, tumbled in value.

For better or for worse, the investment community relies on earnings as its key metric. Stocks are judged not only by their ability to increase earnings quarter over quarter but also by whether they are able to meet or beat a consensus earnings estimate. Like it or not, investors need to keep an eye on consensus numbers to keep tabs on how a stock is likely to perform.