'READING THROUGH THE LINES' PODCAST – TRANSCRIPT

Amanda (00:05)

Welcome to Inside 65. The Reserve Bank of Australia's podcast, where we bring you insights into monetary policy, the financial system, our economy, and the impact of the world at large. From our head office at 65 Martin Place in Sydney, we'll pull together interviews, conversations and explainers as well as include speeches given around Australia to tell you a little bit more about who we are and what we do. If you are interested in more about the Bank, check out our website – rba.gov.au – where you can subscribe to other content, find more information, or contact us directly.

I'm Amanda and today I'm chatting with Callan Windsor, who is a research manager in our Data Science Hub. Callan and fellow data scientist Max Zang have been using a large language model to process and analyse earnings call transcripts to develop new indicators about firms' price-setting behaviours. This research is the subject of one of the articles in our September *Bulletin*, as well as a recent Research Discussions Paper. Earnings call transcripts are a rich and timely source of information about firms' own business conditions, as well as economic and financial conditions more broadly. Callan and Max's research has found the sentiment and language used during these calls provide additional context, nuance and insights that are valuable for economists trying to understand how firms make choices and how those choices affect prices.

Your research sounds quite complex and uses a huge amount of data and a bit of AI wizardry to process it all. Can you give me an overview of what it was you were doing and why?

Callan (01:38)

So, at the outset I might push back a little bit on it being perceived as overly complex, but it is different and so it might appear complex in the sense that we're using a new source of data, namely firms earnings call transcripts, as well as new techniques, so namely large language models. But I do think over the next 10 years or so, folks will be making much more use of text as data and these techniques in natural language processing would be mainstream.

So what we do in this article as we process and analyse the transcripts of earnings calls made by list of Australian firms going back to 2007 using a state-of-the-art large language model. Now we do this to construct new indicators that track how firms are talking and thinking about input costs, demand conditions and final prices with the ultimate aim of trying to tease out some new insights regarding how firms set prices, or at least reinforce existing stylised facts about firms' price setting behaviour.

Now you might be wondering why and what are earnings calls. So, these typically take place a couple of hours after the release of the earnings results by listed companies and during the calls executives provide updates regarding the quantitative results over the previous six months but also explain the context and nuance around which these results came about. And there's also a question-and-answer session that can be quite lengthy and these are really rich exchanges of information that we're able to extract and analyse using these large language models.

Amanda (03:04)

OK, so if I'm following correctly, you're essentially using language processing to ascribe a quantitative measure to something that's essentially qualitative data, and I imagine that removes a lot of the usual human bias in these kinds of assessments. And we use a similar approach when we're analysing the information gathered throughout liaison program. Is that the idea?

Callan (03:26)

Yeah, you're spot on. By using a large language model where to construct indicators from the text of the earnings calls related to our topics of interest – so input costs, demand and final prices – in a way that's kind of consistent and systematic over time. But perhaps more to the point it's just not possible for human to read and analyse the volume of text we're dealing with. In this work we processed and analysed over 5,500 earnings call transcripts and categorised over 750,000 paragraphs of text.

Amanda (03:55)

That's a huge amount of text.

Callan (03:57)

Yeah, and these models can go through them really efficiently and so they're quite complementary to the work you would do as a human analysing these transcripts. And you can actually go wider and deeper. And to your question about the information gathered though liaison program, again, you're right. Information gathered through liaison program at the Bank has been systematically recorded in confidential diary notes and so we wanted to compare our new earnings call indicators to the information that we get from liaison program as a proof of concept. So to do this we collected up all these confidential diary notes that have been written over the past 20 years or so and constructed similar indicators to that that we do for earnings calls and then we compared and contrasted them.

Amanda (04:37)

What were the key things that you discovered?

Callan (04:39)

Well, first what we found is when we aggregated our new earnings call indicators across all firms we found that the indicators closely tracked similar indicators constructed from the RBA's business liaison program as well as from business surveys. And in the case of the business survey indicators, we actually found that our earnings call indicators tended to lead the information that we get from the surveys.

Second, we found that our new indicators also closely track official statistics. For instance, our indicator for final prices closely tracks movements in consumer price inflation. These two things give us confidence in the information we're extracting from the earnings call indicators.

Finally, and probably most interestingly, is when we dig into the firm level indicators we found that price setting intentions appear to change depending on the circumstances that firms face. For instance, firms appear to respond differently to changes in costs versus changes in demand, and according to whether or not costs were increasing or decreasing. We also documented substantial differences across industries. And so we argue that these differences are, worth thinking about when trying to model pricing dynamics in the macro economy.

Amanda (05:47)

Interesting. Was there anything in the research that surprised you?

Callan (05:50)

Perhaps the most surprising thing was just the performance of the large language model that we used – it’s ability to categorise and analyse the text. We basically took an off-the-shelf, what’s called open-source model and used it to classify the earnings call transcripts into our topics of interest without any kind of fine-tuning on our end. And it performed remarkably well, which was surprising.

Second, I was surprised by the nature of the information we were able to extract from the earnings calls. These calls typically discuss results that occurred over the past six months and are delivered with a two-month lag, so my prior was at the information would be mostly backward-looking. But it turns out there's lots of information that are relevant for assessing the current economic conditions. And in the Research Discussion Paper and the *Bulletin* we offer some reasons why this might be the case.

Finally, I'm just constantly surprised by the technical ability of the data scientists that I get to work with. I mean, just collecting these earnings calls was a very large task, let alone analysing them using a modern large language model. And my co-author Max Zang was able to do this very effectively and efficiently. So that's a constant source of almost amazement on my behalf.

Amanda (07:01)

Yeah, it’s a huge… it blows my mind just how much data we are talking about. Thousands upon thousands of paragraphs of text. It’s just quite remarkable. How reliable is this approach? And how do you bring together the different indicators in a meaningful way? You mentioned that they closely track different sources of data that we've got. Is it partly about checking these soft indicators against each other, like you mentioned before, to validate the results to some extent? Is it telling us anything we don't already know?

Callan (07:25)

Yeah, so on the point about reliability, yeah, we do a few things. First, as you say, we do compare our new indices against other soft indicators and our new indicators of track these and sometimes lead them, which is, a good proof of concept. Does this add value? I would argue it does because if we've got two or three sources of independent information telling us the same thing, then we can be confident in the signal we’re extracting from these sources of information. Results consistently obtained from multiple sources of information are harder to dismiss. I think we are adding value in that respect.

We can also check the reliability of these indicators by comparing the new indicators derived from the large language model to simpler methods, for instance derived by just looking up keywords related to our topics of interest. And those two indicators – the complex one that uses a large language model and the simple one that just looks up keywords – do track each other. But what we find is that the accuracy of the large language model-derived indicators is a lot higher. And we're able to assess that by basically just doing spot checks of the classifications. And that's another really, really useful way to check the reliability of your indicators. Basically, just look at 200 classifications, go through 200 paragraphs, and look where the model is performing well and where it might be falling short.

Amanda (08:46)

And the model can learn from that, can’t it?

Callan (08:48)

Exactly. We can adjust what we do from those insights that we get from doing spot checks. For instance, we can change the prompt just a little bit when we're asking the model to characterise the paragraphs into our topics of interest.

Amanda (08:58)

Are there any limitations with this approach?

Callan (09:02)

Yes, for sure. I think more work needs to be done to investigate the nature of the information we're extracting; like actually mapping how information how to flows from earnings calls to liaison information to the news media, and also flows to and from the policy publications that are put out. So how information propagates through these various sources would be interesting and give us more insight about the nature of the information we're extracting from earnings calls.

I think another limitation is constructing indicators that are forward-looking, which might be of more interest. It was actually a really challenging task to categorise the text into past, present and future tense. So that's a limitation and something will continue to work on going forward.

Finally, I think another limitation of the approach we've taken is we're just using an off-the-shelf large language model. I think future work might want to fine-tune the model so it's better suited to analysing earnings calls specifically. But that fine-tuning is quite a computationally intensive task, so you need to have access to the right technology to do that and that's a challenge, especially doing it in a secure environment.

Amanda (10:10)

I'm going to be honest here. There's something that I am a little, I guess you could say, sceptical or unsure about with language processing in that I sometimes wonder how it can handle the nuance of how people use language. And so, what I mean here is that, if I'm understanding it correctly, you need to assign a value and a singular meaning to specific words – like 'crash' might be deemed innately negative or 'soar' might be positive. But a word is not necessarily ways positive or negative and can have different meanings in different contexts. And then if you overlay that with how we use it, so sarcasm or humour, or the fact that earnings calls are to a large extent a way for these firms to communicate to their shareholders and so they have some sort of particular narrative that they want to share, it makes it even more nuanced and complicated. So even just the title – 'Reading through the Lines' – highlights the level of nuance and intelligence needed to assess this kind of data. How do you account for this?

Callan (11:06)

Yeah, you're right to be sceptical. Capturing things like negation, for instance, like 'unemployment increased'. Typically, 'unemployment' is going to be a negative word but 'increased' is typically going to have a positive connotation, so together you might get a neutral outcome. So, how do you capture negation? That's a very challenging task. Also, context. For instance, in a financial stability context the word 'stress-testing' is describing a model and word 'stress' might have a negative connotation.

Amanda (11:32)

Yeah, exactly.

Callan (11:33)

So, yeah, this was a very challenging task in the field of natural language processing. But the current crop of large language models do a very good job of this. And this is because their design, or architecture, enables them to be trained at massive scale. So these models, which often have billions of parameters, are fed massive amounts of textual data and basically tasked with reconstructing corrupted text snippets – for instance, a sentence with a missing word – back into their original form and minimising the distance… or you know, the distance between the correct text snippet and the reconstructed text snippet. And it turns out that if you give these models this task and train them to do it over and over and over again, they found the development understanding of natural language and display actually remarkable and emergent capabilities. And so, this includes the ability to characterise text into topics or themes in a way that is cognisant of the context and nuance of natural language as we do in this work.

Amanda (12:30)

Would this analysis have even been possible for years ago? And what possibilities can you see in the future as these sorts of technologies continue to develop?

Callan (12:38)

No, I think, I mean … A few years ago would have tackled this problem in a very … well, in quite a different way. For example, I probably would have gone about just constructing a very bespoke set of keywords related to the topics of interest and then essentially look them up in the earnings call transcripts and developed an indicator based off the number of hits that I got. So, it's a new world and more broadly, you know, you can see the large increase in this type of work using large language models in just the number of AI-related academic publications. So, you know, that speaks to the fact that this is an emerging technology.

Looking to the future, I see a world where people start working with more fine-tuned versions of these models. So essentially optimising them to perform even better over the specific documents that they're interested in. And finally, I don't think it's inconceivable that in the future large language models will not only be used to kind of analyse digital content but help generate it. So, for instance, by asking questions and analysing the responses almost in real time.

Amanda (12:36)

Great. How do you see the Bank utilising this work in the future?

Callan (12:40)

I think the Bank will continue to use this work as a complement to insights we get from our discussions with businesses, as well as the survey information we track. You might want to be looking at the earnings calls indicators right after earning season and then updating that information between earning season with insights you get from the business surveys and from our liaison program. I think we'll see more work in the future combining the textual indicators from earnings calls with quantitative information that we get and we know how to work with from their balance sheets and income statement data as well. And then finally I see our methodology being used or updated to construct new indicators as well. For instance, you could track how firms are thinking about cyber risk, or climate change, and you can also construct forward or backward-looking indicators as well.

Amanda (14:26)

Callan, thank you for chatting with me today and helping me to better understand your research.

If you want to learn more about Callan and Max's work, head over to our website to read their article in the September edition of our *Bulletin*. You can also find a more technical discussion in their Research Discussion Paper.

Thank you for listening.