Consumer Finance Monitor (Season 3, Episode 1): Using Alternative Data in Credit Underwriting

Speakers: Alan Kaplinsky and Chris Willis

Alan Kaplinsky:

Welcome to the Consumer Finance Monitor Podcast, where we explore important new developments in the world of consumer finance and what they mean for your business, your customers and the industry. I'm your host today, Alan Kaplinsky, the chair of the Consumer Financial Services Group at Ballard Spahr, and I'll be moderating today's program.

Alan Kaplinsky:

For those of you who want even more information, don't forget about our blog which is also called Consumer Finance Monitor. We've been doing the blog since 2011, so there's a lot of content there, including content on the subjects that we're going to be talking about today. We also regularly host webinars on subjects of interest to those of you in the industry. If you want to subscribe to our blog or to get on the list for our webinars, visit us at ballardspahr.com. And if you like our podcast, please let us know about it. Leave us a review wherever you procure your podcasts, be it Apple, Google Play, Spotify or any other podcast site.

Alan Kaplinsky:

So today, I'm very pleased to be joined by Chris Willis, who is the chair of the Consumer Financial Services litigation group, and also the deputy chair of the entire Consumer Financial Services Group. Welcome, Chris.

Chris Willis:

It's great to be here, Alan.

Alan Kaplinsky:

The topic we're going to talk about today is a very recent inner agency statement on alternative data in-credit underwriting. When I say an inner agency statement, I mean one issued by the CFPB and the four bank prudential regulators.

Alan Kaplinsky:

So Chris, we hear all the time about alternative facts, it's on the news all the time. Does this have anything to do with that?

Chris Willis:

No, I don't think it has to do with that. Alternative data really refers to the idea, in the context of this inner agency statement, of using information about individual consumers beyond what you would find on a credit bureau report. We all are familiar with credit bureau reports, they contain other creditors' experiences with a particular consumer about what credit was extended, what the terms were, and what the consumer's repayment performance was over a period of time in the past, and bankruptcy information is also included on most credit reports. But, there are other types of information, things like bank account records, or utility payments, or lease payments, or even public records like vehicle and watercraft registrations, professional licenses, all kind of things like that, even down the things like what you do on the internet, that are available about consumers.

Chris Willis:

So when we speak to alternative data, we're talking about all of those things generally, things that are not on someone's credit bureau report.

Alan Kaplinsky:

Where does the industry get access to all these things? Where are they going to get information about what bank accounts the person has, or all these other things that you listed, Chris? Where's that come from?

Chris Willis:

A lot of them are available from vendors who sell that information to creditors for us in underwriting decisions.

Chris Willis:

For example, there are certain services that will collect consumer's bank account information by consumer's giving the service access to their username and password for their depository bank website, typically. It will scrape information from those websites, and then aggregate it in a way that then can be purchased and used by creditors to look at cash flow information for someone. And likewise, public records are collected from public record sources, like county registrars' offices or things like that. And then, of course information about people's behavior on the internet, which is probably not a great candidate to talk about with respect to credit underwriting because it probably cannot be used for that, consistent with the Fair Credit Reporting Act. But nevertheless, that information is available from entities who monitor consumer's behavior on the web, like social media sites, search engines, things like that.

Alan Kaplinsky:

So Chris, this sounds very futuristic, almost like science fiction. Is the industry actually doing anything with alternative data today, to underwrite loans?

Chris Willis:

Absolutely. It's actually in pretty widespread use already. I think, really, the pioneers in using alternative data first came among fintech lenders, but now even large bank and non-bank financial institutions have put it into use. There are commercial scores that are available from various score providers that incorporate and use alternative data, and these are large national score providers that make that data available to creditors, and they have lots of large creditor subscribers to use these products.

Chris Willis:

Historically, what we've seen is that alternative data has primarily been used to try to allow financial institutions to make credit available to people who don't have very well established credit bureau files, or perhaps no credit bureau file at all. People that we would refer to as thin files or no hits, with the three large national credit bureaus. The first use of alternative data, and the one that I think is still most common today, is to use in underwriting those people who cannot be evaluated using credit bureau data because they don't have enough of it in their own credit bureau files.

Chris Willis:

I think for the future, what's more likely is we might also start to see the use of alternative data to supplement traditional credit bureau data, layer it on top of it, even for people who do have sufficient credit bureau experience to have a credit score. But so far, it's really been primarily for these thin files or no hits, which the CFPB has done three reports about, actually referring to them as credit invisibles.

Alan Kaplinsky:

And, am I right that the CFPB has actually been encouraging the use of alternative data?

Chris Willis:

Yes.

Alan Kaplinsky:

It's something they like.

Chris Willis:

They have. I don't think they're unqualifiedly encouraging it saying, "Do whatever you want and there's no rules," and certainly this inner agency statement underlines that point. But they have, I think, been trying to encourage industry to use it.

Chris Willis:

The things that they've done have included, for example, issuing a no action letter to a fintech lender called Upstart, that was proposing to use alternative data in its underwriting models for approving people for credit transactions. And then, issuing a blog post about the Upstart experience, in which the CFPB stated in its blog, that it had been engaged in monitoring for fair lending risks, the underwriting decisions made by Upstart using its alternative data model. And, that it found that the model was able to increase access to credit without creating any fair lending concerns. When that blog post came out, which was in 2019, it was relatively recent, I really viewed it as a way for the CFPB to publicly encourage others to do what Upstart had done and use this data, in a responsible way, that would increase access to credit without creating fair lending concerns. And basically, illustrating to the public and to the industry that doing so was possible.

Chris Willis:

That's consistent with the informal messaging that we've been seeing from the CFPB over the years. At conferences and in conversations, we've been hearing a similar set of messages from the CFPB's personnel, about the use of alternative data. I think the bureau sees it as a real way to solve the credit invisible problem, and to make access to credit more evenly available to American consumers. And in fact, making credit more available disproportionately advantages those who have thin files or no hits with the credit bureaus, and those tend to fall in certain ethnic or racial groups. So making credit more available to them actually reverses some of the disparate impact that's inherent in our current credit scoring system.

Alan Kaplinsky:

Yeah. Let me ask you this, Chris. Suppose I'm applying for a loan, and I'm a credit invisible. I'm somebody who just graduated from college, I haven't yet established any kind of credit history. I don't have any credit file at the credit bureau. Would I know, when I apply to you for a loan, let's say you're the lender, that you're going to be using alternative data? And, would I know if the answer is yes, what alterative data you're using?

Chris Willis:

You would not know at the time of application, you would just sign an FCRA authorization to allow the lender to obtain information about you from a consumer reporting agency. But after the fact, you would get the required notices under the Fair Credit Reporting Act and Equal Credit Opportunity Act, including for example the credit bureau disclosure that's required by the FCRA. They would state the source of information that was used in the credit decision. If the credit decision was a decline, it would identify the specific aspects of data that led to the decline. So you would find out after the fact, what alternative data was used. At least, from the generic standpoint of what the source of the data was. It certainly would not provide all the detail about all the data elements that were consulted.

Chris Willis:

Just like today, when you apply for credit and a lender uses your credit bureau report from one of the three national agencies, it doesn't list every trade line that was on your credit report. But, it does give you the option of obtaining a copy of your credit report from the provider that was used by the lender, so that would be the way for the consumer to find out what information is available about him or her.

Alan Kaplinsky:

Okay. Let's now get to the inner agency statement that I referenced at the beginning of our podcast. What are the key points of that statement?

Chris Willis:

Well, I think one of the key points to me is that the statement is bland and generic, and didn't tell the industry much that it didn't already know. I view it as an effort by the agencies who issued it to try to encourage the use of alternative data in underwriting, but make it clear that that doesn't come with a pass on applicable consumer protection laws, like fair lending laws, or the Fair Credit Reporting Act, or UDAP principles.

Chris Willis:

What the bulletin actually says is, "We, the agencies who issued it," being the Federal bank regulators and the CFPB, "recognize that alternative data can be useful in making credit more available to underserved populations. And, we expect institutions who decide to use alterative data to do so in a way that is consistent with sound model risk management and consumer protection laws."

Chris Willis:

The agencies go on to highlight the fact, actually in a couple of places, that information about a consumer's bank account activity poses less risk than other forms of alternative data, which they don't specify, might pose. They say this because the bank account data, first of all, relates to a core element of credit underwriting, that is someone's ability to pay based on their cashflow information, so it's intuitively related to someone's ability to repay a credit obligation and is of the nature of information that lenders have considered forever in evaluating a consumer for a loan. It's just this is a better source of information for what they've already considered.

Chris Willis:

They go on to note that the information is fairly transparent, a consumer can understand, "Hey, you're using my bank account information." And, that it comes from a reliable source, that is directly from your checking account records for example. And, for all those reasons, the agencies devote a long paragraph in this relatively short bulletin, that talks about how the use of bank account information is not very risky.

Chris Willis:

The other point that's made in the bulletin is that the agencies consider it to be lower risk to use alternative data in what they refer to as a "second look" model, which is exactly what is described a few minutes ago. That is, taking people who are thin files or no hits, who would be declined under a traditional underwriting model that uses credit bureau data, and using the alternative data to look at them again with an eye toward approving some proportion of them. The CFPB, as I said, has been saying this for years informally, and now all the agencies together have said it in writing. But, they add as a proviso there, just because it's a second look model doesn't exempt it from compliance with consumer protection laws. They were careful to point that out.

Chris Willis:

Those are really the main substantive points of the bulletin. So it's, as I said, a mild encouragement to use alternative data, perhaps a stronger encouragement to use bank account information in particular, and a warning that, just because you're using alternative data to expand access to credit doesn't mean that you don't have to consider things like the Fair Credit Reporting Act and Equal Credit Opportunity Act. The bulletin, honestly, doesn't say that much, and certainly in my view, not much that we didn't already know as industry observers of this issue.

Alan Kaplinsky:

Okay. Let's try to flesh out the inner agencies' statement. Let's talk next about the things that it didn't get into, but which you believe are important for the industry to be aware of. And, you mention a couple of things, Fair Credit Reporting Act, and fair lending, which I assume you're referring to the Equal Credit Opportunity Act.

Alan Kaplinsky:

Let's start with the Fair Credit Reporting Act. What would the concerns be there?

Chris Willis:

Well, the Fair Credit Reporting Act will, if you get information about a consumer that is used for determining the consumer's eligibility for a credit product, there's a good chance that whatever you have obtained will be deemed a consumer report under the Fair Credit Reporting Act, and subject to the Act's various requirements.

Chris Willis:

Those requirements include measures taken to ensure its accuracy and integrity, and a dispute mechanism. In other words, the consumer has to have a mechanism to dispute what is in his or her consumer report. And then, of course it comes with obligations on the lender using it, to give adverse action notices and credit bureau disclosures, of the nature that I was talking about before.

Chris Willis:

So the concern with alternative data is that, when you use it for underwriting, which is one of the specific permissible purposes under the Fair Credit Reporting Act, that you make sure that everybody knows that it is a consumer report, and everybody's complying, both the furnishers of the information, the entity supplying it to the lender and the user of it, the lender itself, that it is a consumer report and that all of the requirements of the FCRA, at every stage of the information lifecycle, are being adhered to. If you don't do those things, if any of those three things is missing, then someone may be in violation of the Fair Credit Reporting Act, and obviously we want to stay away from that, and the agencies are warning everybody about that in this bulletin.

Alan Kaplinsky:

Okay. Now, let's turn to fair lending and the ECOA. What are the concerns there, Chris?

Chris Willis:

Well, they are the same kind of concerns that we could have with any kind of scoring model, that either the model could incorporate something that is a protected characteristic. So for example, a piece of alternative data could be your gender, or your race or ethnicity, or your religion, or your familiar status, or something like that. So to the extent that a protected characteristic were considered in a scoring model, that would pose a pretty high likelihood of violating the Equal Credit Opportunity Act, or the Fair Housing Act for mortgage credit, and state analogs to that statute.

Chris Willis:

Then we have to be concerned, if we get those out of the way and we don't have those in our model, that we don't have anything that's a close proxy for one of those. If there are things like zip code, for example, the Federal banking agencies, and the CFPB, and the FFIEC, inner agency fair lending exam procedures, the appendix, warn lenders about using zip code. But, zip code is a piece of alternative data even though it could be predictive of credit risk and repayment performance, it's so strongly correlated with race and ethnicity that it's deemed to be a proxy for those and therefore have a high likelihood of violating the Equal Credit Opportunity Act if it's used in an underwriting model. Our concerns are, first and foremost, the use of protected characteristics or close proxies, in the alternative data that we're considering.

Chris Willis:

Then, if we get past that, we have to worry about disparate impact. That is, does the operation of the model impose a disparate impact on some particular protected class basis? Gender, age, race, ethnicity, something like that. And, if it does, does the use of that data fail to advance the predictiveness of the model in a way that would provide a defense for the lender in using that data? This really just comes down to sound model development and risk management processes, where you are thoroughly testing each variable for predictiveness, looking for alternative variables, and using only the most predictive ones in the model that can be shown to be empirically derived and statistically sound as predictive of credit risk, in this example, for an underwriting model. We have that general disparate impact risk, which is similar to the same analysis we'd do for a traditional credit bureau model.

Chris Willis:

But, the difference is that traditional credit bureau models are built on variables that the regulators have accepted the use of for years. Whether you have delinquencies on prior loans, or a prior bankruptcy, or what your credit limit was or your credit utilization was, all that stuff on a credit bureau report has been used without objection in credit scoring models for decades. Now, we may have new elements on information that don't enjoy that decades of experience with the Federal banking regulators and the CFPB, and therefore we are less sure about how they will react to it. I think that means that institutions have to be extra careful about first excluding characteristics and close proxies, and then showing the predictiveness and lack of alternatives for other data elements that they use in their models.

Alan Kaplinsky:

Well Chris, if you don't mind, let me drill down on some of the things that you said. Because, I'm wondering as a practical matter, whether lenders that have decided to use alternative data and they go to some consumer reporting agency that specializes in that area, and I assume it's going to be an agency other than the big three, though I don't know. Is the big three dealing with alternative data right now?

Chris Willis:

I'm not aware of the extent to which they're doing it. The ones I was thinking of are not the big three. There are large providers that provide alternative data scores, but it's not the big three.

Alan Kaplinsky:

Okay. So you get a call from a client today that says to you, "Chris, we're going to start using alternative data because we think it will really help us increase the volume of loans that we're going to be able to make to people who have a thin or no credit file at one of the bureaus. I don't really know yet, what is in this credit score that will be given to me for each applicant. What I do know is that I've talked to some other people in the industry and they've said this is a reliable company. Is there anything more I need to do, Chris? What should I be doing here?"

Chris Willis:

Yeah. If your question is directed to the fair lending side of it, I think the best due diligence on utilizing something like that would be to find out what are the data elements that are included in the score. You don't necessarily have to know what weight they have because the score provider may be reluctant to provide that, but knowing what's in there, and having your own opportunity to see is it something that looks like it may be offensive to a regulator.

Chris Willis:

The second thing that you can ask for is any fair lending analyses of the score product you're buying, that have been done by the score provider. One of the score providers that I'm aware of, that does sell these to large lenders, has a very comprehensive fair lending study that it has done of its particular score product that shows not only does it not have a disparate impact, but it disproportionately makes credit accessible to African-American and Hispanic consumers who are not able to get credit using credit bureau data. If the score provider you're dealing with has something like that, and you feel that

the study is credible by looking at it in detail, that could also provide a lot of comfort for you in using the score. And, a lot of documentation to establish to your regulator that you did your due diligence on the score before deciding to use it.

Alan Kaplinsky:

Okay. We talked about Fair Credit Reporting Act and ECOA. Are there any other legal issues, either Federal or state, that you have to be sensitive to, such as UDAP?

Chris Willis:

Well, UDAP is mentioned in the inner agency bulletin, as I mentioned earlier in the podcast. There, with respect to UDAP principles, it really refers, I think, in the agencies' minds, to the idea of both transparency to the consumer. In other words, does the consumer understand what the data is and how it relates to the credit decision that's being made? And, it also relates to the issue of has the consumer gotten clear notice that the data is being collected, and shared and used, for credit decisions.

Chris Willis:

This is more of a privacy issue, but was a disclosure made of, "Hey, you're giving us access to your bank account data. We're going to store that, and when you apply for a loan we're going to make that available to lenders who are evaluating you for credit." You could imagine a disclosure that says that clearly and prominently, and that would be fine from a UDAP standpoint. Or, you could imagine a failure of disclosure in that regard, which could raise UDAP considerations. That transparency and privacy issue is, I think, one part of the UDAP analysis.

Chris Willis:

The other one is really just to ask, how intuitively related to a credit decision is this data? There, we go back to this discussion in the inner agency statement that we're talking about today. It talks about bank account data being highly intuitively related to a credit decision, because it directly shows you what the customer's cash flow is. But, if you took some crazy data element that seemed totally unrelated to a credit transaction, it would be more difficult to establish it as fair from the standpoint of a UDAP analysis to use in a credit decision. And certainly, you'd have to have some extremely compelling data to show that it does accurately predict repayment risk. I think this intuitive relationship of the data and evaluating your credit application is another element of the UDAP analysis.

Alan Kaplinsky:

Yeah, okay. You mentioned earlier in your comments, Chris, that social media could be a type of alternative data that might be utilized. But, you also cautioned that that probably would not be a good idea for somebody to use that as part of the credit underwriting decisions. I'd like to know why is it that you made that comment?

Chris Willis:

Yeah. It's really mostly having to do with the Fair Credit Reporting Act, but also these UDAP principles that you and I were just talking about.

Chris Willis:

When information is collected about you, in terms of what you like on social media or who your friends are, or things like that, then that data collection, if it is then sold to someone, is not really meeting the standards of accuracy and integrity under the Fair Credit Reporting Act, and the provider of that data does not think it is a consumer reporting agency governed by the FCRA.

Chris Willis:

There's also no dispute process. Is there a process to dispute what you liked on social media? "No, I didn't like that." Or, what you looked at, or what you responded to, or what affinity groups you belong to or whatever like that. There is not accuracy measurement in place, there's no dispute mechanism in place, so the data's not compliant with the Fair Credit Reporting Act and therefore, couldn't be used for a credit underwriting decision.

Chris Willis:

And then, on top of that, even if you then ingested it and used it for an underwriting decision, you'd have to put that stuff on your adverse action notice and that is discussed in the inner agency statement, too. It reminds lenders about adverse action notices. I haven't yet encountered a business person in the consumer lending industry that says, "I want to put on my adverse action notices that I declined you because you liked X on Facebook." Or, "You shopped for Y on Amazon," or something like that. There's a great reluctance to put stuff like that on adverse action notices.

Chris Willis:

And then, of course there's the UDAP issue, which we were just talking about. There's not a real good intuitive relationship between stuff you do on social media, or what you shopped for on the internet, and what your payment performance should be on a credit obligation. Maybe there is a connection there in the data, but it's not an intuitive connection that a regulator will be pre-disposed to accept. So you raise UDAP risk there, too.

Chris Willis:

But first and foremost, I think it's just off the table because it doesn't comply with the Fair Credit Reporting Act. And the entities that supply that information don't want to be consumer reporting agencies, so they would resist any use of their data that would turn them into one.

Alan Kaplinsky:

Okay, got it. Chris, before we bring this to a conclusion, are there any closing remarks that you'd like to make about this subject?

Chris Willis:

Yes. What I would say is I think the current regulatory environment is one that is favorable for the expanded use of alternative data in credit underwriting. I think the regulators, as I said, are trying to encourage lenders to use it, and I think that the primary worry that a lot of lenders have is that, with these untested data elements, and I mean untested in terms of regulatory reaction, that they're taking some high degree of fair lending risk if the use of those data elements ends up having a disparate impact on the basis of some protected characteristic under the Equal Credit Opportunity Act. I don't view that risk to be very high, based purely on disparate impact.

Chris Willis:

Again, we've screened off protected characteristics and close proxies for our model. If we have what's left over, I think if you have a good model development process that shows the predictive power of the elements that are being included in the model, and shows that there are not available alternatives that are more predictive, then I think you can put an alternative data model squarely into the language in Reg B that says it is fine for a creditor to use an empirically derived, statistically sound scoring model, and that that kind of model should be subject to the same kind of conclusion, from a fair lending standpoint, as one that uses credit bureau elements.

Chris Willis:

I think there's a lot of reluctance to use alternative data because of what I consider to be an overblown fear on the disparate impact side, and I don't see any evidence to support that fear. I feel like lenders are really cheating themselves out of a significant business advantage because of a regulatory concern that I don't think has much substance to it. You still have to do

this in a smart way, and take into account the fair lending issues that we've been talking about here in detail, but it is totally doable. And, I just sense a lot of reluctance to do it out of an inflated fear of the fair lending reaction under the disparate impact theory.

Chris Willis:

My own closing thought would be we should try to rid ourselves of that mental obstacle and move forward with what this data can enable us to do, both for the good of consumers and for our lending institutions.

Alan Kaplinsky:

Let me ask you one final question. And that is, what is the relationship, if any, between using alternative data for credit underwriting decisions and also using artificial intelligence, or AI?

Chris Willis:

Yeah, the relationship is that they typically, but do not require ... They don't have to go together, but they frequently do. Because artificial intelligence and machine learning algorithms do a better job of ingesting large amounts of data, hundreds or perhaps thousands of variables, figuring out which ones are the most predictive and when they are predictive of what in combination with one another, so you typically see the models go hand-in-hand with the machine learning algorithms using alternative data. But, you could use alternative data using a traditional logistic regression algorithm, just like we've used for decades with credit bureau data. Likewise, you could use a machine learning algorithm on traditional credit bureau data and make a more accurate, better model that way.

Chris Willis:

There's no hard connection between the two, it's just as a matter of practice they frequently go hand-in-hand.

Alan Kaplinsky:

Okay. Well Chris, thank you very much for sharing your insight and wisdom on a very important subject, and a subject that is likely to become of increasing importance in the coming years.

Alan Kaplinsky:

Also, let me thank all of our listeners who downloaded and listened to today's podcast. Let me remind you again to visit our website, www.ballardspahr.com, where you can subscribe to our show in Apple Podcasts, Google Play, Spotify or any other podcast platform. And don't forget to check out our blog, consumerfinancemonitor.com, for daily insights of the consumer finance industry. If you have any questions or suggestions for our podcast show, please email them to us at podcast@ballardspahr.com. Our podcast generally gets released on a weekly basis, typically on a Thursday except during certain holiday periods. Thank you, once again.