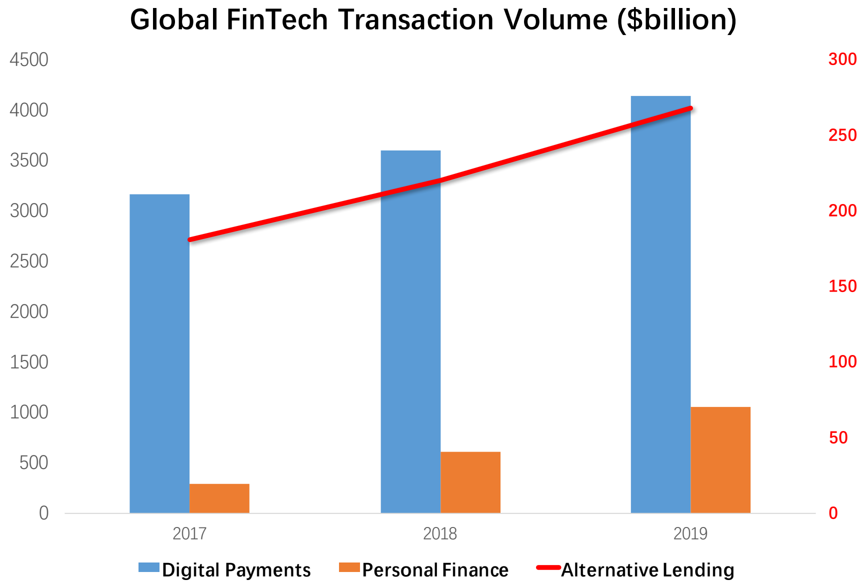
The Value of Privacy: Evidence from Online Borrowers

Huan Tang

London School of Economics

Privacy concerns matter for FinTech

- **FinTech is growing fast**
  - Global VC investment in FinTech 2018: $112 billion
  - Global transaction volume 2019: $5.5 trillion



Privacy concerns matter for FinTech

- **FinTech is growing fast**
  - Global VC investment in FinTech 2018: $112 billion
  - Global transaction volume 2019: $5.5 trillion

- **Data is crucial for FinTech**, e.g.,
  - Credit allocation: *Kabbage* (valued at $1.2bn) uses social media patterns
  - Data sales: *Credit Karma* ($4bn) generates revenue from customer referrals

- **But data is not "free"**
  - Individuals may be reluctant to share private data
  - Firm revenue is constrained by privacy concerns

Q: How big are those constraints for FinTech?

A: Value of privacy

Privacy concerns matter for FinTech

- **FinTech is growing fast**
  - Global VC investment in FinTech 2018: \$112 billion
  - Global transaction volume 2019: \$5.5 trillion

- **Data is crucial for FinTech**, e.g.,
  - Credit allocation: *Kabbage* (valued at \$1.2bn) uses social media patterns
  - Data sales: *Credit Karma* (\$4bn) generates revenue from customer referrals

- **But data is not "free"**
  - Individuals may be reluctant to share private data
  - Firm revenue is constrained by privacy concerns

- **Privacy paradox**
  - People keep relinquishing personal data, e.g., Facebook and Google
  - Evidence from lab experiments suggests small/zero value of privacy

Research question

Do people value privacy? How much?

Implications for FinTech firms?

Setting: Online lending in China

- **Empirical challenge**:
  - Measurement of demand for/willingness to pay for privacy

Setting: Online lending in China

- **Empirical challenge**:
  - Measurement of demand for/willingness to pay for privacy
- **Why online lending in China?**
  - Borrowers face a tradeoff: privacy vs. credit access
    - No credit bureau
    - Borrowers underserved by banks
  - \$218bn market size, > 50% global market share

Setting: Online lending in China

- **Empirical challenge**:
  - Measurement of demand for/willingness to pay for privacy
- **Why online lending in China?**
  - Borrowers face a tradeoff: privacy vs. credit access
    - No credit bureau
    - Borrowers underserved by banks
  - \$218bn market size, > 50% global market share
- **Large-scale field experiments** (315,785 participants)
  - Vary disclosure requirements and loan terms

Setting: Online lending in China

- **Empirical challenge**:
  - Measurement of demand for/willingness to pay for privacy
- **Why online lending in China?**
  - Borrowers face a tradeoff: privacy vs. credit access
    - No credit bureau
    - Borrowers underserved by banks
  - $218bn market size, > 50% global market share
- **Large-scale field experiments** (315,785 participants)
  - Vary disclosure requirements and loan terms
- **A structural model**
  - Value of privacy
  - Borrower welfare
  - Platform profit

Key findings

- **Social network ID and employer contact valued at:**
  - 230 RMB ($33) > half-day salary
  - 8% of the value of a foregone loan

Key findings

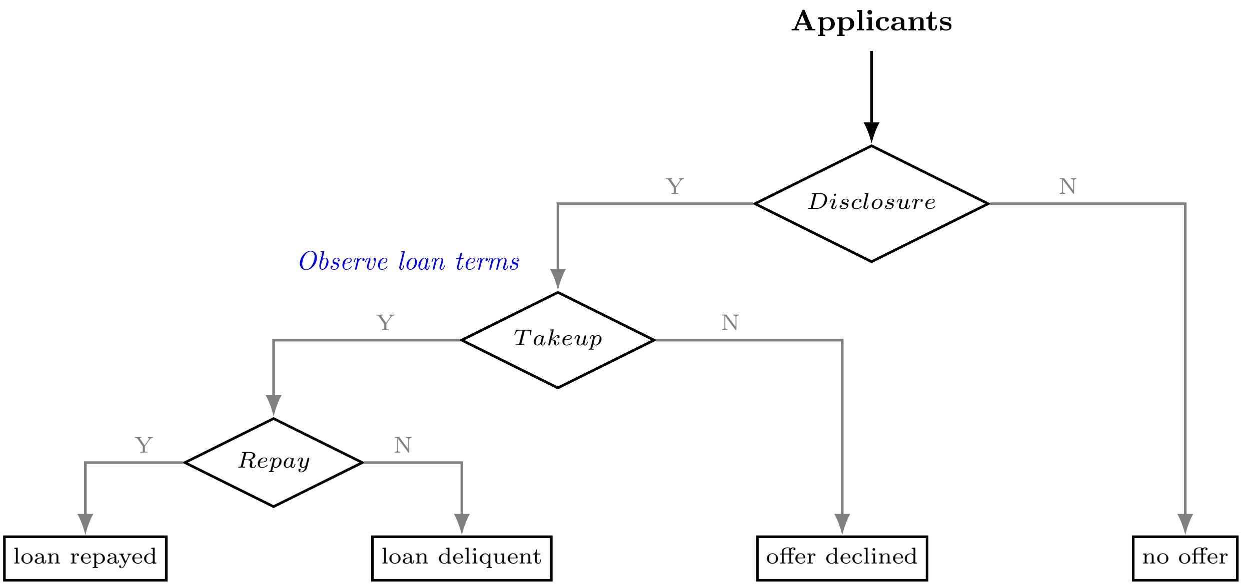
- **Social network ID and employer contact valued at:**
  - 230 RMB ($33) > half-day salary
  - 8% of the value of a foregone loan
- **Possible mechanisms**
  - An incentive to hide negative info ✗
  - Exertion ✗
  - Intrinsic preferences for privacy ✓

Key findings

- **Social network ID and employer contact valued at:**
  - 230 RMB (\$33) > half-day salary
  - 8% of the value of a foregone loan
- **Possible mechanisms**
  - An incentive to hide negative info ✗
  - Exertion ✗
  - Intrinsic preferences for privacy ✓
- **Structural model: cost of data collection**
  - Borrower welfare ↓13%
  - Platform expected revenue per applicant ↓10%

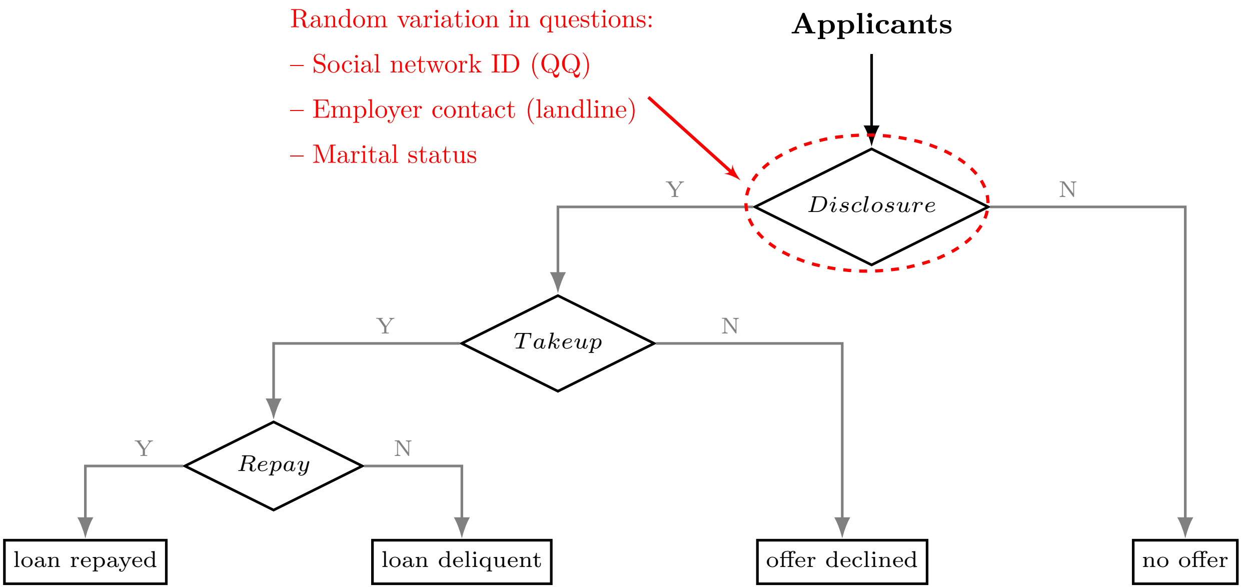
Borrower Decision Process

- **A typical loan**: 3,770 RMB ($540), amortizing, 12 months, 11% interest rate, 29% fee, 15% delinquency rate



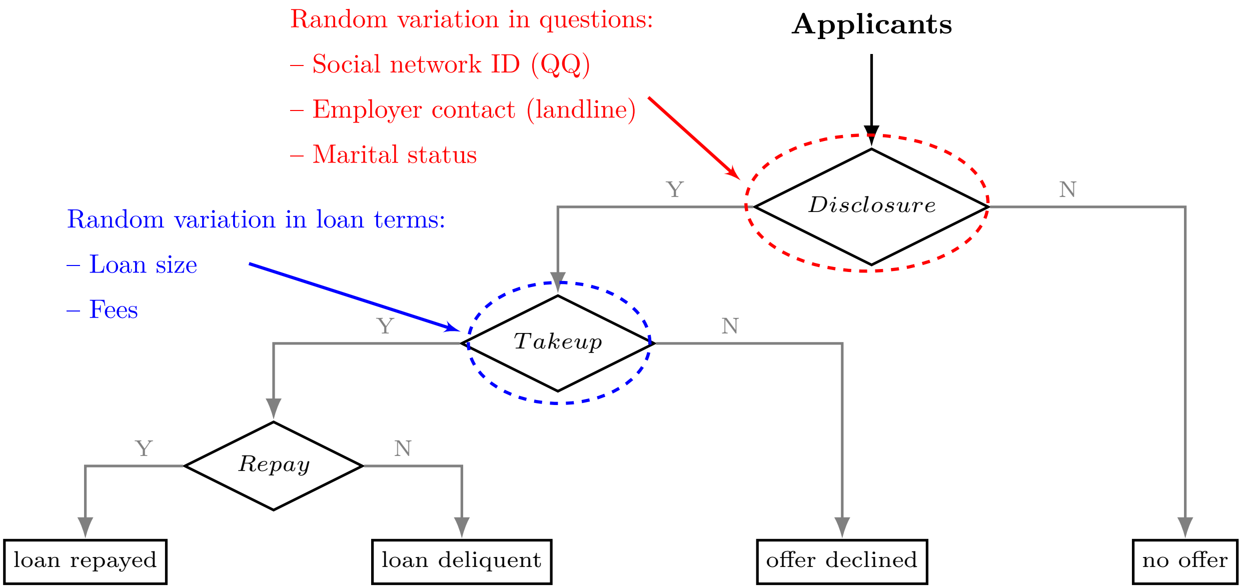
Borrower Decision Process

- **A typical loan**: 3,770 RMB ($540), amortizing, 12 months, 11% interest rate, 29% fee, 15% delinquency rate



Borrower Decision Process

- **A typical loan**: 3,770 RMB ($540), amortizing, 12 months, 11% interest rate, 29% fee, 15% delinquency rate



Do applicants value privacy?

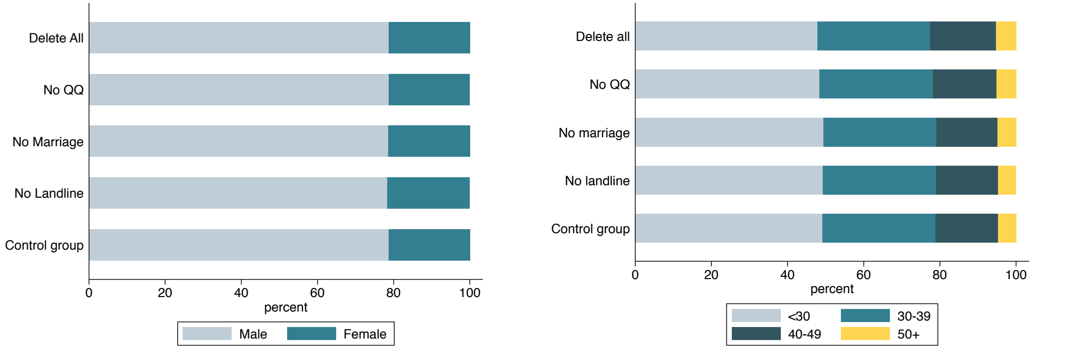
the disclosure RCT

The disclosure RCT: set-up

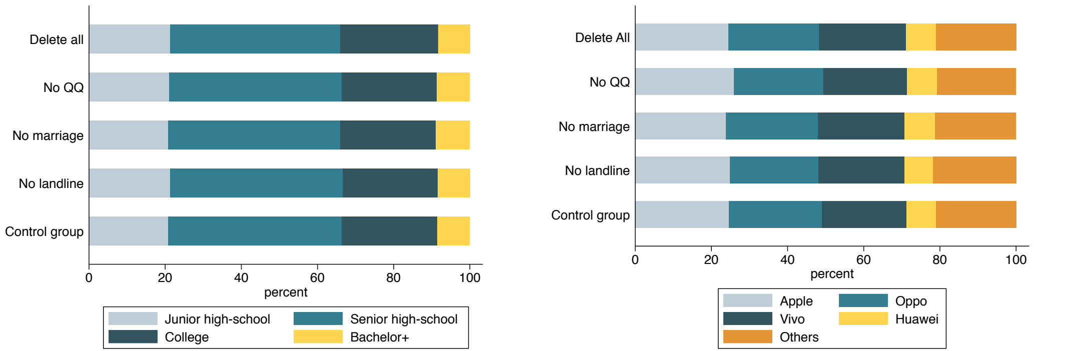
- **270,388 first-time applicants**
- **4 treatments**:
  - social network ID (QQ ID $\simeq$ Whatsapp + email) “no QQ”
  - marital status “no marriage”
  - employer contact (landline phone number) “no landline”
  - all three items “delete all”

Quality of randomization

(a) Gender (b) Age



(c) Education (d) Mobile



Treatment effects of disclosure requirements

- **Applicants are reluctant to disclose:** QQ & landline

	270,388 first-time applicants	Completion%
control	QQ → Marriage → Landline	**30.76**
no QQ	~~QQ~~ → Marriage → Landline	+1.04***

Treatment effects of disclosure requirements

- **Applicants are reluctant to disclose:** QQ & landline

	270,388 first-time applicants	Completion%
control	QQ → Marriage → Landline	**30.76**
no QQ	~~QQ~~ → Marriage → Landline	+1.04***
no marriage	QQ → ~~Marriage~~ → Landline	+0.04

Treatment effects of disclosure requirements

- **Applicants are reluctant to disclose:** QQ & landline

	270,388 first-time applicants			Completion%
control	QQ →	Marriage →	Landline	**30.76**
no QQ	QQ →	Marriage →	Landline	+1.04***
no marriage	QQ →	Marriage →	Landline	+0.04
no landline	QQ →	Marriage →	Landline	+0.61***

Treatment effects of disclosure requirements

- **Applicants are reluctant to disclose:** QQ & landline

	270,388 first-time applicants			Completion%
control	QQ →	Marriage →	Landline	**30.76**
no QQ	QQ →	Marriage →	Landline	+1.04***
no marriage	QQ →	Marriage →	Landline	+0.04
no landline	QQ →	Marriage →	Landline	+0.61***
delete all	QQ →	Marriage →	Landline	+1.28***

What explains completion rates?

- An incentive to hide negative information ✗ table
  - info. used in pricing
  - info. used in debt collection
- Exertion ✗ table
- Difficulty in recollecting information ✗ table
- An intrinsic preference for privacy ✓ table
  - *Female* and *old* attach higher value to privacy
  - No heterogeneity across *income* and *education*
  - Goldfarb and Tucker (2012), Prince and Wallsten (2020)

next section

What is the monetary value of privacy?

the loan RCT

The loan RCT: set-up

- **46,170 borrowers who have completed application questionnaire**
- **Two treatments**:
  - loan size $\times$ 2
  - fee reduction (1/2 fee $\approx$ \$128)

	46,170 applicants			take-up rate
control	regular loan			**57.6%**
treatment 1	loan size×2			**+6.5%**
treatment 2	loan size×2	+	fee reduction (\$128)	**+11.9%**

Back-of-the-Envelope

- **Loan demand = disclosure × take-up**

disclosure requirement fee reduction

- **QQ and landline** $= \$128 \times \frac{\mathbf{1.28}}{\mathbf{5.44}} = \mathbf{\$30}$

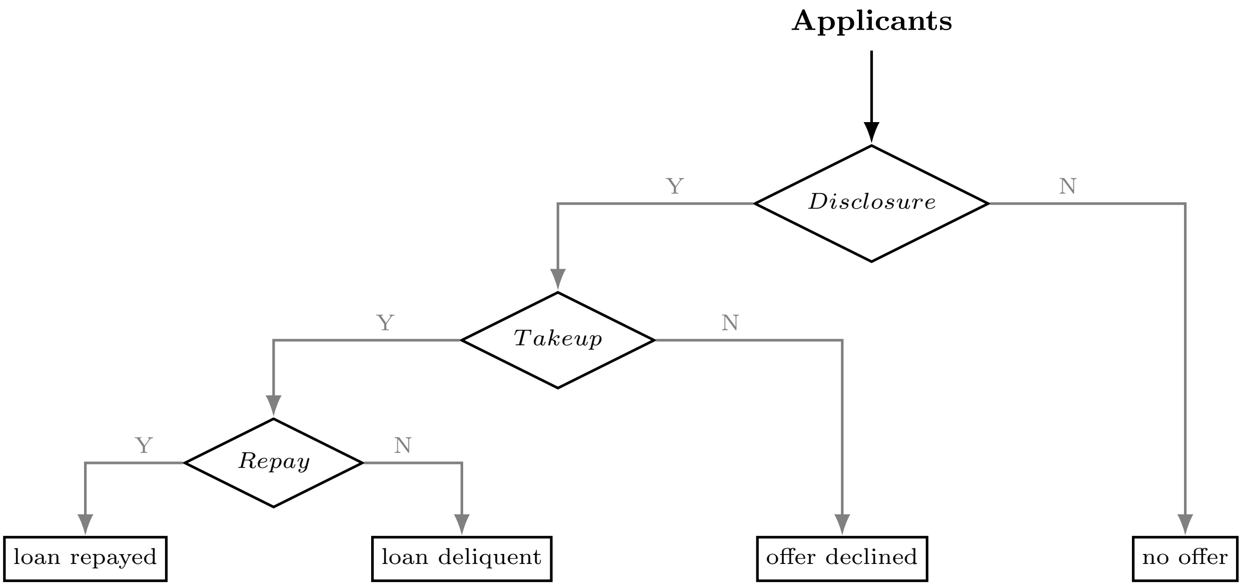
	Δ **demand%**	[t]
	disclosure	
remove QQ + landline	**1.28**	[4.91]
	take-up	
fee reduction ($128)	**5.44**	[6.84]

Why A Structural Approach

1. **Potential selection on sensitivity to loan price:**
   - disclosing applicants $\neq$ exiting applicants
2. **No insights on borrower welfare or platform profit**
   - Need a welfare measure
   - Need a model for firm profit

A Structural Model for Privacy Demand

Borrower Decision Process



Note: Individual make *Disclosure* decision before observing loan terms

Linking the model to the data

Individuals' choices are given by:

$$\mathbf{D1} = \mathbb{1}\{X'\gamma_{\mathcal{X}} - \theta_{1,qq}\mathbb{1}_{qq} - \theta_{1,marr}\mathbb{1}_{marr} + \gamma_{1,\mathcal{L}}E_L - \gamma_{1,\mathcal{R}}E_R + \varepsilon^{D1} \geqslant 0\}$$

$$\mathbf{D2} = \mathbb{1}\{X'\gamma_{\mathcal{X}} - \theta_{2,qq}\mathbb{1}_{qq} - \theta_{2,marr}\mathbb{1}_{marr} - \theta_{2,ll}\mathbb{1}_{ll} + \gamma_{2,\mathcal{L}}E_L - \gamma_{2,\mathcal{R}}E_R + \varepsilon^{D2} \geqslant 0\}$$

$$\mathbf{T} = \mathbb{1}\{X'\beta_{\mathcal{X}} + \beta_{\mathcal{L}}L_i - \beta_{\mathcal{R}}R_i + \sum \beta_q\mathbb{1}_q + \varepsilon^{T} \geqslant 0\}$$

$$\mathbf{F} = \mathbb{1}\{X'\alpha_{\mathcal{X}} + \alpha_{\mathcal{R}}R_i + \sum \alpha_q\mathbb{1}_q + \varepsilon^{F} \geqslant 0\}$$

- **Key coefficients:**
  1. **Value of privacy:** $\frac{\theta_{1,qq}}{\gamma_{1,\mathcal{L}}}$, $\frac{\theta_{1,marr}}{\gamma_{1,\mathcal{L}}}$, $\frac{\theta_{2,ll}}{\gamma_{2,\mathcal{L}}}$
  2. **Selection**:
     - Unobservable: $\varepsilon_i^D, \varepsilon_i^T, \varepsilon_i^F$

Estimation Results

Demand Estimates

- **Requiring QQ and landline *decreases* disclosure probability**
- **(Expected) larger loans and lower fees *increase* disclosure probability**

	D - page 1	t	**D - page 2**	t
QQ	-0.023	(-14.20)	0.016	(6.33)
marriage	-0.002	(-1.03)	0.003	(1.46)
landline	-	-	-0.007	(-3.23)
loan (000s)	0.203	(16.03)	0.083	(9.01)
repayment (000s)	-0.139	(-14.53)	-0.054	(-7.63)

$$\left.\begin{aligned} \textbf{QQ}: \frac{0.023}{0.203} \times 1000 = 145 \text{ RMB} = \$21 \\ \textbf{landline}: \frac{0.007}{0.083} \times 1000 = 85 \text{ RMB} = \$12 \end{aligned}\right\} > \textbf{half-day salary}$$

Value of loans

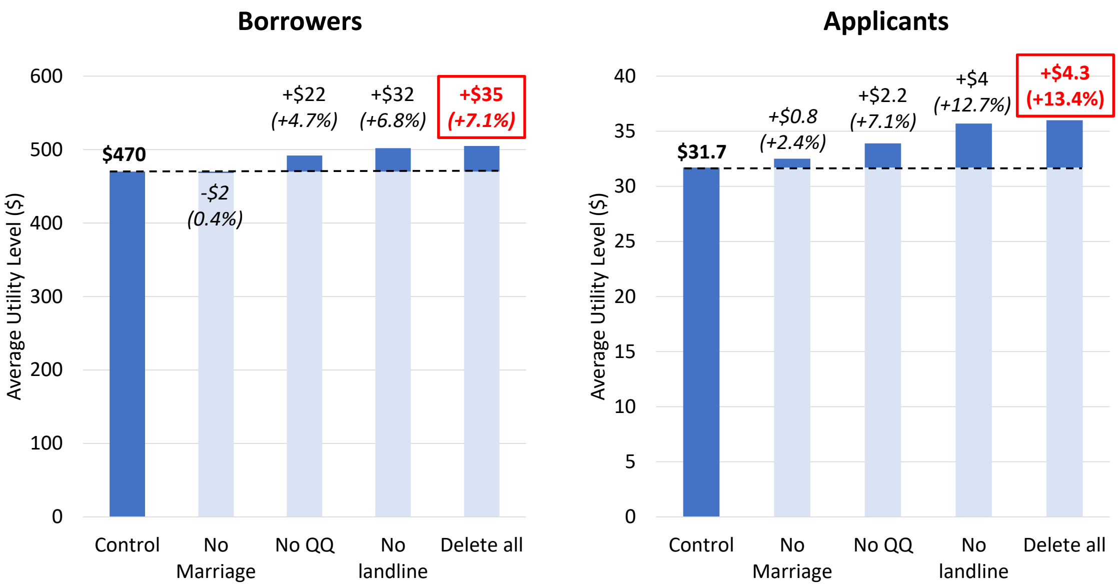
- **Average present value of a loan**: $\hat{V}(loan) := L - \hat{\delta}R = \$420$ [a]

- **QQ + landline: $\mathbf{\$33/\$420 = 7.8\%}$**

[a] Estimated annual discount factor $\hat{\delta} = 0.44$: borrowers are liquidity constrained

Borrower welfare

- **Utility in monetary terms:** $V_T/\hat{\gamma}_L$
- **Avg. utility of successful borrowers** $\uparrow$ **7.4%** - *intensive margin*
- **Avg. utility of all applicants** $\uparrow$ **13.4%** - *intensive + extensive margins*



Platform Profit

Implications for platform profit

- **Data is not "free"**
  - Appropriate data elicitation incentive, *or*
  - Lower loan demand
- **Platform profit depends on data collection policy:**

Step 1: predict demand and repayment using demand estimates ✓

Step 2: calculate platform revenue ✓

Step 3: cost of lending is such that under the current loan terms L & R, firm profit is maximized

Implications for platform profit

- **Data is not "free"**
  - Appropriate data elicitation incentive, *or*
  - Lower loan demand
- **Platform profit depends on data collection policy:**

Step 1: predict demand and repayment using demand estimates ✓

Step 2: calculate platform revenue ✓

Step 3: cost of lending is such that under the current loan terms L & R, firm profit is maximized

⇒11 cents per dollar originated

Counterfactual: expected revenue per applicant

- **Collecting QQ and landline decreases platform profit by 10%**

	No question	**With questions**	**Difference**
delete all	**4.88**	**4.40**	**-9.9%**

Conclusion

- **Individuals attach positive value to privacy**
- **Data collection could lead to a deadweight loss**
- **A generalizable methodology for firms and regulators**

Is it driven by an incentive to hide negative info.?

- **No difference in loan performance or loan grade**

$$risk = X'\beta_1 + \sum_{j=1}^{5} \gamma_j \; group_j + \varepsilon$$

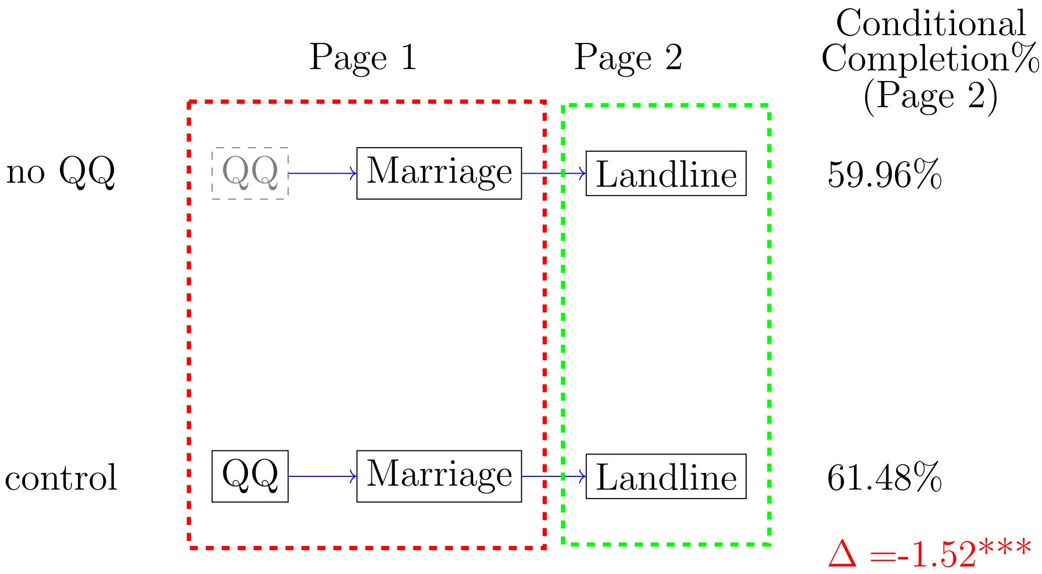
	grade	pre-approval	loan size	fee	fraction of payments
control	4.12***	0.43***	3797.65***	28.58***	0.85***
	(447.21)	(110.89)	(94.63)	(469.01)	(157.56)
delete all	0.02	-0.01	21.30	0.01	0.00
	(1.50)	(-1.64)	(0.38)	(0.12)	(0.35)
no QQ	-0.00	-0.01**	29.22	-0.09	0.01
	(-0.13)	(-2.14)	(0.52)	(-1.00)	(0.85)
no marriage	-0.02	-0.00	39.15	-0.01	-0.00
	(-1.58)	(-0.72)	(0.69)	(-0.15)	(-0.23)
no landline	0.00	-0.00	16.51	-0.12	0.01
	(0.28)	(-0.18)	(0.29)	(-1.34)	(1.51)
Observations	73,051	73,051	31,006	30,992	15,532
R^2	0.00	0.00	0.00	0.00	0.00

*Einav, Finkelstein, Cullen (2010) QJE

Go Back

Is it driven by exertion?

- **Applicants in the control group are NOT more tired on page 2**



◂ Go Back

Is it driven by memory?

- **Only employed applicants are reluctant to disclose landline**

$$Completion = \beta_0 + \boldsymbol{\beta_1}\ no\ landline + \varepsilon$$

	complete page 2	
	employed	**unemployed**
no landline	0.02***	-0.01
	(2.75)	(-1.43)
Constant	0.53***	0.63***
	(79.61)	(103.32)
Observations	21,477	20,862
R^2	0.02	0.01

Go Back

An intrinsic preference for privacy

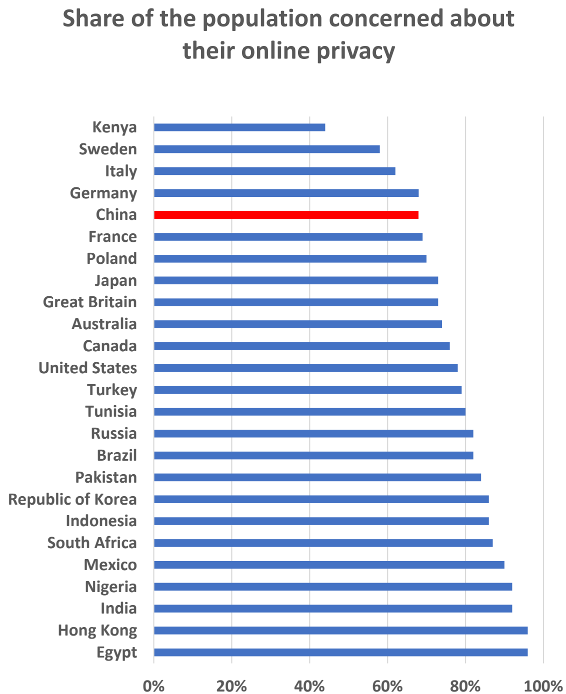
- ***Female*** and ***old*** people less willing to disclose personal data
- No heterogeneity across *income* and *education*
- Consistent with Goldfarb and Tucker (2012), Prince and Wallsten (2020)

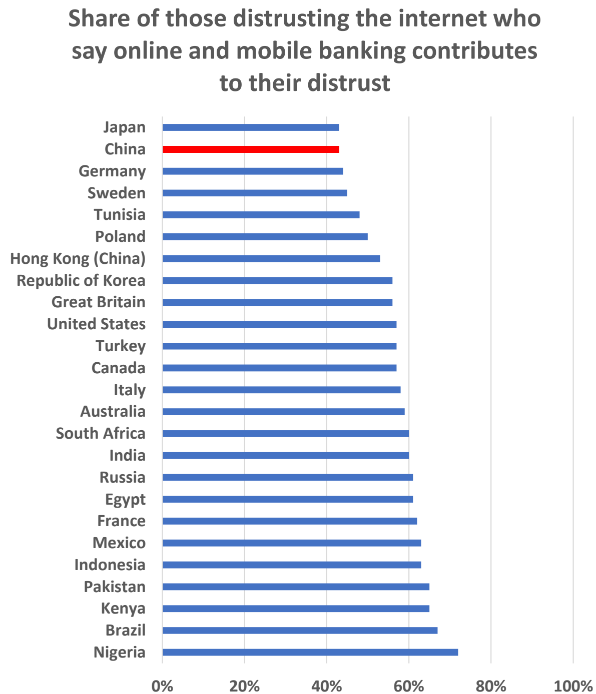
$$Completion = \beta_0\ treatment + \beta_1 X + \boldsymbol{\beta_2}\ treatment * X + \varepsilon$$

Treatment	**no QQ** complete page 1		**no landline** complete page 2	
treatment	0.01**	0.01	0.01	0.01
	(2.56)	(1.42)	(1.50)	(1.35)
female	0.04***		0.04***	
	(6.90)		(5.79)	
treatment × female	0.02**		-0.00	
	(2.15)		(-0.28)	
old		-0.02***		-0.07***
		(-4.22)		(-13.65)
treatment × old		0.01*		-0.00
		(1.93)		(-0.35)
Observations	71,956	71,956	69,986	69,413
R^2	0.00	0.00	0.00	0.00

Go Back

Privacy Concerns Across Countries





Source: CIGI-IPSOS Survey – Internet Security & Trust, 2019

Go Back