Addressing Remitting Behavior Using An Ordinal Classification Approach

IWINAC'13

5th. International Work-conference on the Interplay between Natural and Artificial Computation

P. Campoy-Muñoz

Universidad Loyola Andalucía

P.A. Gutiérrez & **C. Hervás Martínez**

AYRNA Research Group: http://www.uco.es/ayrna

Universidad de Córdoba

June 12, 2013

OUTLINE

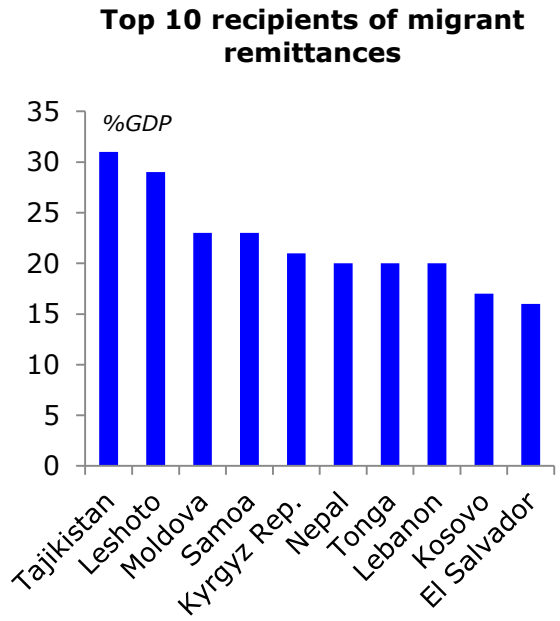
1. THE PROBLEM
2. OUR PROPOSAL
3. EXPERIMENTAL DESIGN
4. RESULTS
5. CONCLUSSIONS

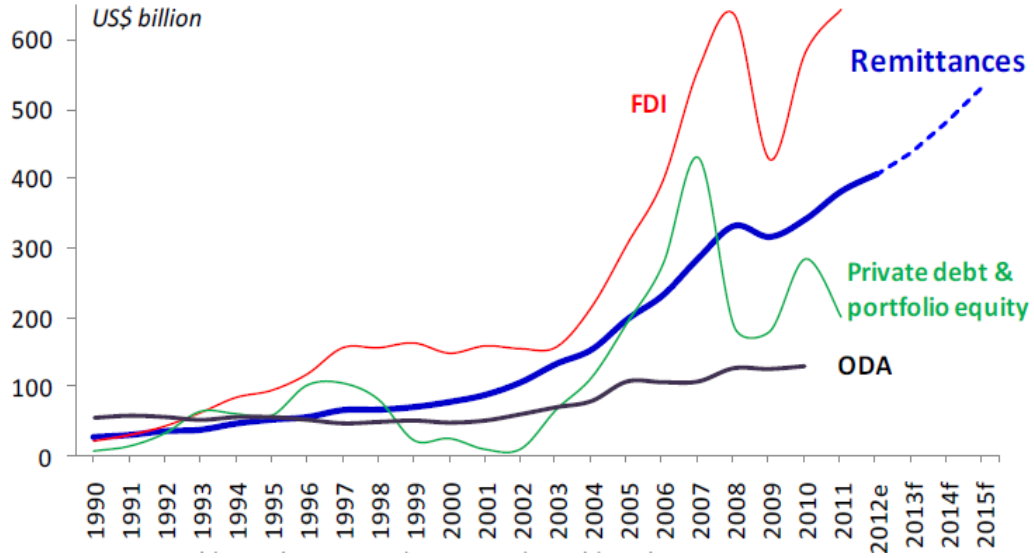
THE PROBLEM

REMITTANCES: transfers made by migrants who are employed and have lived, at least one year, in other economies.

2011

Worldwide	**215 MM immigrants**	⇔	**501.000 MM US$**
Developing countries	**171 MM immigrants**	⇔	**372.000 MM US$**





Source: World Bank Indicators, 2011

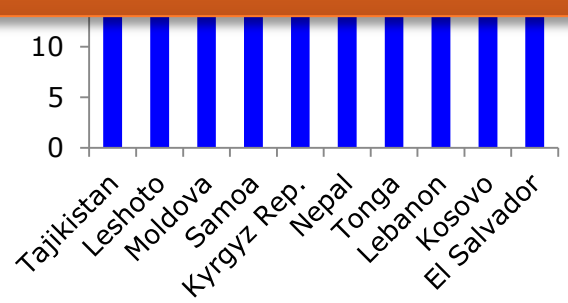
THE PROBLEM

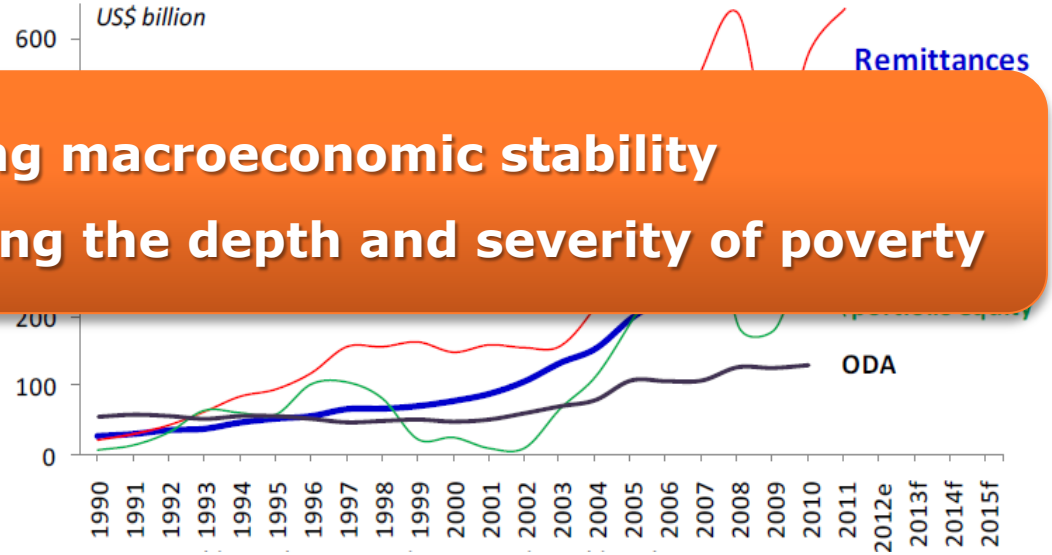
REMITTANCES: transfers made by migrants who are employed and have lived, at least one year, in other economies.

2011

Worldwide	**215 MM immigrants**	⇔	**501.000 MM US$**
Developing countries	**171 MM immigrants**	⇔	**372.000 MM US$**

Top 10 recipients of migrant remittances





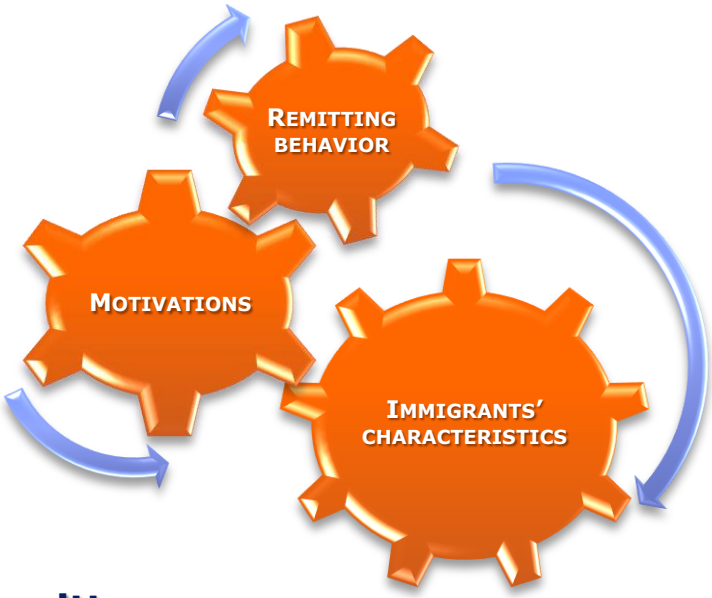
Country level→ Improving macroeconomic stability

Household level→Reducing the depth and severity of poverty

Source: World Bank Indicators, 2011

The Problem

Why do some migrants send much (less) money than others?



$$f : X \rightarrow Y$$

$$x \in X \subseteq R^k$$ **Immigrants' characteristics**

Target variable

$$y \in Y = \{C_1, C_2, \ldots, C_Q\}$$ **Remittance levels**

Traditional approach ➡ **Ordinal regression** (logit and probit models)

- Learn a rule to predict categories or labels in an ordinal scale.
- The labels are discrete, but there is a natural order among them.

Performance** can be **hampered** by the assigning of numerical values because the **distance** between **classes** are **unknown

OUR PROPOSAL

CLASSIFICATION APPROACH → **Machine Learning Methods**

- An algorithm extracts information given a group of label data in order to be able of predicting the label of new unseen data

$$f : X \rightarrow Y$$

$$x \in X \subseteq R^k$$ **Immigrants' characteristics**

TARGET VARIABLE

$$y \in Y = \{C_1 \prec C_2 \prec \ldots \prec C_Q\}$$ **Remittance levels**

ORDINAL CLASSIFIERS

Remitting behaviour would be an **ordinal outcome**, as other economic behaviors (saving, investment, credit consume…)

TARGET VARIABLE

$$y \in Y = \{C_1, C_2, \ldots, C_Q\}$$ **Remittance levels**

NOMINAL CLASSIFIERS

A **posteriori definition** of the target variable makes its application suitable

Our Proposal

Ordinal Classifiers ➡ **Support Vector Machine Approach**

- Trying to find the optimal pararell separating hyperplane $f(\mathbf{w},\mathbf{x},b)=\mathbf{w}\cdot\mathbf{x}+b$
- To construct the decision function $y_i=\text{sign}(f(\mathbf{w},\mathbf{x},b))$, $x_i \in X$, $y_i \in Y\{-1,1\}$
- The introduction of slack-variables ξ_i relax the hard-margin constrain, allowing to deal with non-separable sets

BINARY DECOMPOSITION

Support Vector Machine with Ordered Partitions (SVMOP)

- Q-1 binary SVM classifiers
- Consider the ordinal nature by comparison between the label of a pattern and a given rank k→ y_{qi}= -1, $\forall q < k$ and y_{qi}= + 1,$\forall q \geq k$
- Explicit weights are imposed over the patterns
- Decision of the different binary classifiers were combined by using associated probabilities, $p_q=P(y \succ C_q \,|\mathbf{x})$, $1 \leq q \leq Q$-1.

Our Proposal

Ordinal Classifiers ➡ **Support Vector Machine Approach**

TRESHOLD MODELS

Support Vector Machine for Ordinal Regression with Explicit Constraint (SVOREX)

- Assuming the ordinal response variable is the indicator of an unobserved continuous variable
- A function $f(\mathbf{x})$ that predicts the real-valued outcomes
- A threshold vector $\mathbf{b} \in R^{J-1}$ to represent the intervals in the range of $f(\mathbf{x})$, where $b_1 \leq b_2 \leq \ldots \leq b_{J-1}$.
- The SVOREX parallel separating hyperplanes with the same normal vector $\mathbf{w}$ but different thresholds b_j.

Our Proposal

Performance evaluation

CONTINGENCY OR CONFUSION MATRIX $M(f)$ in a classification problem with Q classes and N training or testing patterns:

$$M(f)=\left\{n_{ij};\sum_{ij=1}^{Q}n_{ij}=N\right\}$$

where n_{ij} represents the number of times the patterns are predicted by classier *f* to be in class *j* when they really belong to class *i* .

Metrics

MAX $$CCR=\frac{1}{N}\sum_{i=1}^{Q}I(y_i^*=y_i)$$

Correct Clasification Rate ⇨*Percentage of immigrants for which the remittance level (class) is correctly predicted*

MAX $$S_q=\frac{n_{qq}}{n_q}$$

Sensitivity of *q*-th class ⇨ *Percentage of immigrants well classified within each class*

MIN $$MAE=\frac{1}{N}\sum_{i=1}^{N}\left|O(y_i)-O(y_i^*)\right|$$

Mean Absolute Error⇨ *Deviation of the predicted class with respect to the real one*

Experimental Design

Dataset description

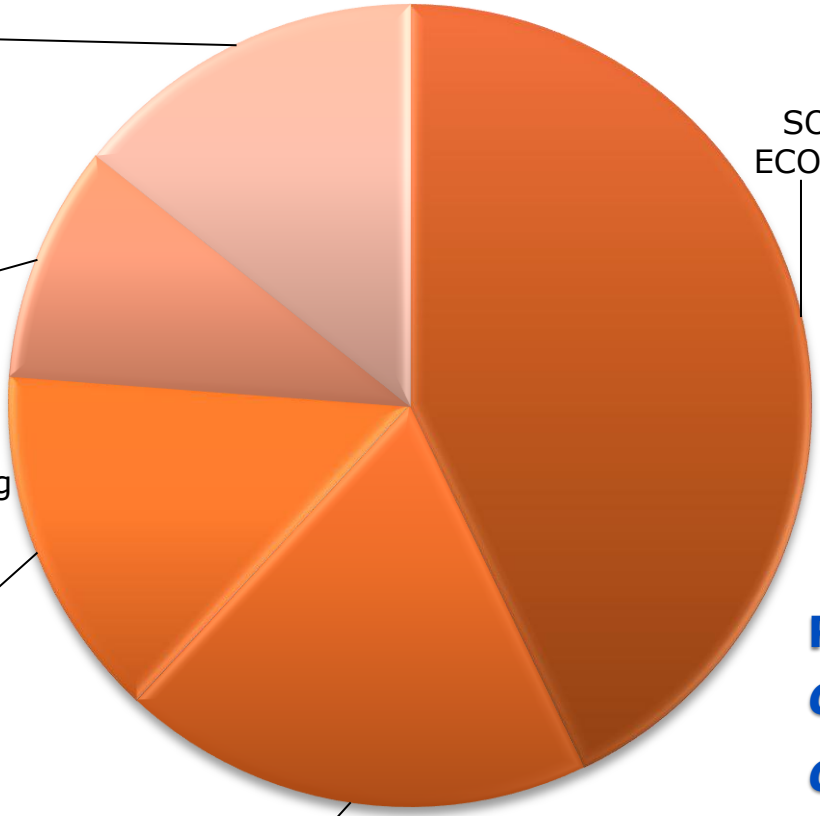
National Immigration Survey (NIS-2007) carried out by Spanish Statistical Institute

- 15,465 foreing borns, aged 16 or over and residing in Spain for at least 1 year
- About 250 variables cover several issues of immigrants' life in Spain.

Sample evidence:

- Ecuadorian immigrants (**599** individuals), one of the largest collective
- **21 variables** related to remitting behaviour, grouped into **5 sets**: socioeconomic variables, labor market situation, linkages with home country (HC),migratory strategy and remittances information.
- **Target variable**: **the amount of remittances sent overseas during the last year**

Experimental Design



REMITTANCES INFORMATION
- Recipients
- Modes
- Frequency

MIGRATORY STRATEGY
- Migratory plan *5y*
- Family reagrouping

LINKAGES HC
- Contact
- Visits
- Ownership

LABOR MARKET
- Monthly income
- Sector of activity
- Occupation
- Length of the contract

SOCIO-ECONOMIC
- Age
- Gender
- Marital Status
- #Children
- Household composition
- Education
- Length of stay
- Legal status
- Ownership At Spain

Remittance Levels (y_i)

C_1: $0 \leq y_i < 250$ €

C_2: $250 \leq y_i < 1{,}718$ €

C_3: $y_i \geq 1{,}718$ €

Experimental Design

		CLASIFFIERS	SOFTWARE
Logistic-based	Nominal	MLogistic	WEKA
		SLogistic	WEKA
Artificial Neural Networks		Multilayer Perceptron Network (**MLP**)	WEKA
		Radial Basis Function Network (**RBF**)	WEKA
Support Vector Machine		Support Vector Machine (**SVC1v1**)	MATLAB
	Ordinal	Support Vector Machine with Ordered Partitions (**SVMOP**)	MATLAB
		Support Vector Machine for Ordinal Regression with Explicit Constraint **SVOREX**)	MATLAB

- 10-fold cross-validation
- Default configuration of parameter setting

- 10-fold cross-validation
- Nested 5-fold cross-validation (over training set) to adjust hyperparameter values $C \in \{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$; $\sigma \in \{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$

RESULTS

ordinal classifiers are better than nominal ones for determining the migrant belonging to the remittances classes

the cost of misclassification is lower for ordinal classifiers compared to nominal ones

ordinal classifiers yields more balanced outcomes by class than nominal ones

Variable set	Classifiers	CCR	$S_{CLASS\ 1}$	$S_{CLASS\ 2}$	$S_{CLASS\ 3}$	MAE
Socio-economic variables	MLogistic	47,41%	43,92%	62,03%	9,08%	65,95%
	Slogistic	40,05%	46,52%	42,92%	14,07%	76,97%
	MLP	40,72%	43,92%	42,58%	15,08%	74,66%
	RBF	46,41%	39,09%	61,10%	8,62%	67,29%
	SVC1V1	44,74%	33,95%	41,16%	57,27%	68,29%
	SVMOP	45,74%	11,64%	64,39%	58,77%	61,94%
	SVOREX	44,90%	17,46%	76,87%	40,13%	59,44%
Labor market situation	MLogistic	50,75%	48,77%	61,47%	14,83%	62,44%
	Slogistic	35,89%	35,44%	47,60%	11,11%	82,32%
	MLP	37,38%	64,06%	30,45%	19,74%	86,68%
	RBF	51,42%	47,69%	65,63%	13,38%	60,94%
	SVC1V1	48,56%	46,61%	37,13%	60,48%	64,97%
	SVMOP	46,25%	16,43%	57,79%	61,97%	62,44%
	SVOREX	47,41%	34,53%	63,39%	44,35%	60,27%
Linkages with HC	MLogistic	52,75%	50,91%	65,65%	11,09%	59,77%
	Slogistic	43,24%	42,34%	53,55%	8,84%	71,12%
	MLP	38,72%	66,17%	30,97%	18,15%	84,34%
	RBF	52,60%	48,77%	68,44%	11,42%	59,93%
	SVC1V1	51,75%	44,50%	40,21%	68,42%	60,61%
	SVMOP	48,92%	31,75%	48,45%	64,29%	61,77%
	SVOREX	47,42%	33,89%	63,34%	44,91%	59,93%
Migratory strategy	MLogistic	55,42%	53,98%	66,56%	9,75%	55,78%
	Slogistic	42,07%	42,84%	53,61%	8,50%	72,29%
	MLP	41,55%	80,91%	30,50%	7,04%	81,50%
	RBF	54,42%	51,81%	68,83%	10,66%	57,44%
	SVC1V1	54,24%	58,10%	37,53%	65,63%	56,95%
	SVMOP	52,91%	50,76%	38,66%	67,49%	57,77%
	SVOREX	50,74%	51,75%	43,37%	56,45%	58,11%
Remittaces information	MLogistic	66,77%	79,39%	64,65%	8,50%	36,74%
	Slogistic	62,61%	*80,96%*	62,03%	14,62%	40,39%
	MLP	56,58%	**87,87%**	52,25%	6,59%	51,94%
	RBF	69,29%	76,73%	**70,26%**	8,36%	32,54%
	SVC1V1	69,95%	75,67%	59,24%	**74,50%**	32,05%
	SVMOP	**71,13%**	77,78%	65,97%	*69,91%*	**30,87%**
	SVOREX	*69,96%*	78,30%	*66,84%*	65,26%	*31,37%*

RESULTS

IMMIGRANTS' PROFILES BASED ON SUPPORT VECTORS OF SVMOP CLASSIFIER

C_1

Young individual, secondary education or below

Living with his/her spouse, at least one parents and sometimes one of her children

Longest stay in Spain and higher percentage with legal status

Employed into construction or service sectors (elementary occupations)

Preference for remaining in Spain but not reagrouping

Remit sometime to parents and children by agency services and to a larger extent only occasionally

C_2

A young woman with secondary studies

Living with her spouse, one of her parents and sometimes any one of her children

Employed in services sector and household activities

She mainly sent money back to her parents and to a lesser extent to her children monthly, quarterly or occasionally, especially agency services but also post office service

C_3

Older man with more than two children

Living with his spouse and to a lesser extent with other family members

He has spent shorter time in Spain compared to the other two groups, but his monthly earnings are the highest

He monthly remits to his parents, as in previous cases, but mostly to his children by using official channels

CONCLUSSIONS

- ✓ *The ordinal classifiers allows to predict reasonably well the remittance level according to immigrants' characteristics*

- ✓ *Tackling the problem as an ordinal classification task can provide accurate information of the remitting patterns among immigrants*

  The observed profiles gives us information that can be used to design measures oriented to promote remitting behaviour among immigrants. For example, to facilitate investment at home country for immigrants belonging to class C1

Addressing Remitting Behavior Using An Ordinal Classification Approach

THANK YOU FOR YOUR ATTENTION!

P. Campoy-Muñoz

Universidad Loyola Andalucía

P.A. Gutiérrez & **C. Hervás Martínez**

AYRNA Research Group: http://www.uco.es/ayrna

Universidad de Córdoba