

Effets marginaux dans les modèles logit à effets fixes

Xavier D'Haultfœuille (CREST-ENSAE)

travail joint avec Laurent Davezies (CREST-ENSAE)
et Louise Laage (Georgetown University)

JMS

Mars 2022

Outline

- 1 Introduction
- 2 Identification
- 3 Estimation de Δ
- 4 Simulations
- 5 Conclusion

Effets causaux dans les modèles binaires de panel

- Pour l'individu i , on observe $(X_{it}, Y_{it})_{t=1, \dots, T}$ avec $X_{it} = (X_{it1}, \dots, X_{itp}) \in \mathbb{R}^p$ et $Y_{it} \in \{0, 1\}$, une variable dépendante binaire.
- Effet de X_{itk} sur Y_{it} ?
- Considérons le modèle de variables potentielles :

$$Y_{it}(x_t) = f(x_t, \alpha_i, \varepsilon_{it})$$

et supposons qu'on observe $Y_{it} = Y_{it}(X_{it})$.

- Problème : α_i , inobservée, est a priori corrélée aux X_{it} .
- On suppose en revanche les $(\varepsilon_{i1}, \dots, \varepsilon_{iT})$ indépendants de (X_{i1}, \dots, X_{iT}) .
- Notons $F(x_t, a) := P(Y_{it} = 1 | X_{it} = x_t, \alpha_i = a)$.

Paramètres d'intérêt usuels

- Pour simplifier, omettons i et considérons les effets à la dernière période.
- Si X_{Tk} continu, un paramètre usuel est l'effet marginal moyen (=AME) :

$$\Delta := E \left[\frac{\partial F}{\partial x_k}(X_{Tk}, \alpha) \right].$$

- AME=effet sur Y_T d'un chgt universel, exogène et infinitésimal de X_{Tk} .
- Si $X_{Tk} \in \{0, 1\}$, un paramètre usuel est l'effet moyen du traitement (=ATE) :

$$\Delta^{ATE} := E \left[F(X_T^{(1)}, \alpha) - F(X_T^{(0)}, \alpha) \right],$$

où $X_T^{(j)} = X_T$ sauf sur la coordonnée k , où X_{Tk} est remplacé par $j \in \{0, 1\}$.

- ATE=effet sur Y_T d'un changement universel et exogène de X_{Tk} de 0 à 1.
- Si $\alpha \perp\!\!\!\perp X_T$, on a $\Delta = E [\partial E(Y_T|X_T)/\partial x_k]$. Sinon ?

La pratique habituelle

- Suivant Angrist (2001) et Angrist & Pischke (2008), les économistes appliqués supposent très souvent F linéaire :

$$F(X_t, \alpha) = X_t' \beta_0 + \alpha.$$

- Alors Δ (ou Δ^{ATE}) = k -ième coordonnée de

$$\beta_0 = E[(X_T - X_{T-1})(X_T - X_{T-1})']^{-1} E[(X_T - X_{T-1})(Y_T - Y_{T-1})].$$

- Idée derrière : même si cette hypothèse est fausse, ce que nous récupérons est la meilleure approximation linéaire de la vraie F .
- Pourtant, les résultats peuvent être trompeurs pour au moins deux raisons :
 - les estimateurs within ou de diff. 1ère n'utilisent que les "movers" ; or, les "stayers", pour qui $X_1 = \dots = X_T$, peuvent être très \neq ;
 - les non-linéarités peuvent encore avoir de l'importance \Rightarrow la meilleure approx. linéaire peut même conduire à se tromper de signe sur Δ !

Une alternative : le modèle logit à effets fixes

- Modèle logit avec effets fixes :

$$Y_t = \mathbb{1}\{X_t' \beta_0 + \alpha + \varepsilon_t \geq 0\}$$

$$\varepsilon_t | X, \alpha \sim \text{logistique, i.i.d. selon } t \leq T. \quad (1)$$

- Rappel : fonction de répartition logistique $\Lambda(u) := 1/(1 + \exp(-u))$.
- T est fixe (panel “court”).
- Approche “effets fixes” : la distribution de $\alpha | X$ (avec $X := (X_1', \dots, X_T')$) n'est pas contrainte.
- Avantages de l'utilisation de ce modèle pour les données de résultats binaires :
 1. Le modèle permet des effets marginaux / de traitement hétérogènes ;
 2. Le modèle tient compte de $E(Y_t | X, \alpha) \in (0, 1)$.
- Mais négligé car aucune méthode pour obtenir Δ ou Δ^{ATE} jusqu'à présent.

Contribution du papier

- Nous montrons que nous pouvons estimer l'AME et l'ATE très simplement dans ce modèle.
- Nous montrons d'abord que l'on peut obtenir simplement des bornes minimales théoriques sur ces quantités.
- Sur la base de cette analyse, nous suggérons deux voies pour l'estimation :
 1. L'estimation des bornes minimales.
(Nécessite une estimation non paramétrique et, pour l'inférence, des conditions de régularité sur $F_{\alpha|X}$)
 2. L'estimation d'une approximation de Δ et du biais correspondant.
(Pas optimale mais très simple et fonctionne très bien en pratique)
- Notre analyse s'étend à d'autres paramètres et modèles similaires.

Lien avec la littérature

Outline

- 1 Introduction
- 2 Identification**
- 3 Estimation de Δ
- 4 Simulations
- 5 Conclusion

Identification et estimation de β_0

- Estimation efficace de β_0 déjà considérée par Rasch (1961) ; voir également Andersen (1970) et Chamberlain (1980).
- Idée : $S := \sum_{t=1}^T Y_t$ est une statistique suffisante pour α :

$$Y := (Y_1, \dots, Y_T) \perp\!\!\!\perp \alpha \mid X, S.$$

⇒ Utiliser la vraisemblance conditionnelle pour identifier β_0 :

$$\beta_0 = \arg \max_{\beta} E [\ln \mathcal{L}_c(Y|X, S; \beta)]$$

avec $\mathcal{L}_c(y|x, s; \beta) = P(Y = y|X = x, S = s)$.

- On estime alors β_0 par :

$$\hat{\beta} = \arg \max_{\beta} \sum_{i=1}^n \ln \mathcal{L}_c(Y_i|X_i, S_i; \beta)$$

- Cet estimateur est asymptotiquement normal (et as. optimal).

AME et AME conditionnel

- Note : on se concentre sur Δ , cf. le papier pour l'ATE.
- Δ vérifie

$$\Delta = \beta_{0k} E[\Lambda'(X_T' \beta_0 + \alpha)].$$

- Ainsi $\Delta = E[\Delta(X)]$, avec $X = (X_1', \dots, X_T)'$ et

$$\begin{aligned} \Delta(x) &:= \beta_{0k} E[\Lambda'(x_T' \beta_0 + \alpha) | X = x] \\ &= \beta_{0k} \int \Lambda'(x_T' \beta_0 + a) dF_{\alpha|X}(a|x). \end{aligned}$$

- On se concentre tout d'abord sur $\Delta(x)$.

Expression de l'AME conditionnel

- Définissons, pour $t = 0, \dots, T$,

$$c_t(x) = E \left[\mathbb{1}\{S \geq t\} \binom{T-t}{S-t} \frac{\exp(Sx'_T \beta_0)}{C_S(x; \beta_0)} \mid X = x \right],$$

$$\sum_{t=0}^{T+1} \lambda_t(x) u^t = u(1-u) \prod_{t=1}^{T-1} (1 + u(\exp((x_t - x_T)' \beta_0) - 1)).$$

Lemme principal

Supposons (1) vérifié, $E[\sum_{(s,t)} (X_s - X_t)(X_s - X_t)']$ inversible et notons \mathcal{D} l'ensemble des mesures de probabilité sur $[0; 1]$. Alors :

$$\Delta(x) = \beta_{0k} \left[\sum_{t=1}^T \lambda_t(x) c_t(x) + \lambda_{T+1}(x) c_0(x) q_T(x) \right],$$

où $q_T(x)$ appartient à l'ensemble :

$$\left\{ q : \exists \mu \in \mathcal{D} : q = \int_0^1 u^{T+1} d\mu(u), \int_0^1 u^t d\mu(u) = \frac{c_t(x)}{c_0(x)}, t = 1, \dots, T \right\}.$$

Bornes sur $q_T(x)$ et l'AME conditionnel

- On peut alors calculer des bornes (optimales) sur $q_T(x)$ puis sur $\Delta(x)$.
- En effet, si l'on connaît les T premiers moments d'une distribution sur $[0, 1]$, on peut simplement calculer les bornes optimales sur le moment d'ordre $T + 1$.
- Aucune optimisation n'est requise, on utilise simplement des déterminants de Hankel \Rightarrow calcul immédiat.
- On peut aussi montrer par exemple que

$$\overline{\Delta}(x) - \underline{\Delta}(x) \leq \frac{|\beta_{0k}|}{2^{T+1}} \prod_{t=1}^{T-1} \frac{|\exp(x'_t \beta_0) - \exp(x'_T \beta_0)|}{\exp(x'_t \beta_0) + \exp(x'_T \beta_0)} \leq \frac{|\beta_{0k}|}{2^{T+1}}.$$

\Rightarrow Bornes très informatives en général.

Détails sur les dét. de Hankel

Outline

- 1 Introduction
- 2 Identification
- 3 Estimation de Δ**
- 4 Simulations
- 5 Conclusion

1ère idée : résumé

- Idée : utiliser des estimateurs “plug-in” basés sur la stratégie d'identification.
- Il en résulte des estimateurs convergents des bornes de Δ .
- Nous montrons que les estimateurs convergent en racine- n et sont asymptotiquement normaux (en général) sous des conditions de régularité.
- Intérêt : approche optimale asymptotiquement.
- Inconvénient : nécessite une estimation non paramétrique, donc :
 1. un choix de paramètres d'ajustement ;
 2. une possible imprécision si $pT = \dim(X)$ est grand.

2ème idée : résumé

- Le lemme principal montre que

$$\Delta = \beta_{0k} E \left[\sum_{t=0}^S \frac{\lambda_t(X) \binom{T-t}{S-t} \exp(SX'_T \beta_0)}{C_S(X, \beta_0)} + c_0(X) \lambda_{T+1}(X) \int_0^1 u^{T+1} d\mu_X(u) \right]$$

pour un certain $\mu_X \in \mathcal{D}$.

- Δ n'est pas identifié simplement à cause de $\int_0^1 u^{T+1} d\mu_X(u)$.
- En particulier, tous les termes $\int_0^1 u^k d\mu_X(u)$, $k \leq T$ sont connus :

$$\int_0^1 u^k d\mu_X(u) = c_k(X)/c_0(X).$$

- Idée : utiliser une bonne approximation de $u \mapsto u^{T+1}$ par un polynôme de degré T .

2ème idée : résumé

- En pratique on cherche :

$$b^* = \operatorname{argmin}_{b \in \mathbb{R}^{T+1}} \sup_{u \in [0,1]} \left| u^{T+1} - \sum_{k=0}^T b_k u^k \right| \quad (2)$$

- On considère alors l'estimation de

$$\tilde{\Delta} = \beta_{0k} E \left[\sum_{t=0}^S \frac{a_t(X) \binom{T-t}{S-t} \exp(SX'_T \beta_0)}{C_S(X, \beta_0)} \right],$$

avec $a_t(x) = \lambda_t(X) + b_t^* \lambda_{T+1}(X)$ pour $t \in \{0, \dots, T\}$.

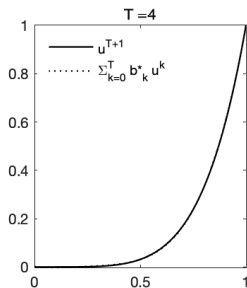
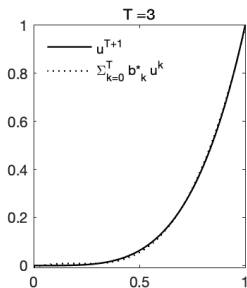
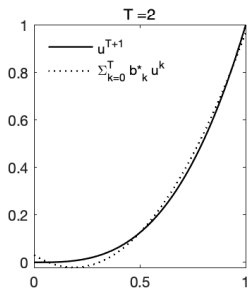
- L'estimateur est biaisé pour Δ ...
- ... Mais on peut estimer une borne sur $|\tilde{\Delta} - \Delta|$ et ainsi faire de l'inférence valide sur Δ !

Calcul et estimation de $\tilde{\Delta}$

- (2) a été résolu par Tchebychev sur $[-1, 1]$. Solution : polynôme de Tchebychev de première espèce (\mathbb{T}_{T+1}).
- Solution sur $[0, 1]$: $\mathbb{T}_{T+1}(u) = 2^{-(T+1)}\mathbb{T}_{T+1}(2u - 1)$.

$\Rightarrow -b^* =$ coefficients de \mathbb{T}_{T+1} , hors le coeff. du terme dominant.

- Approximation déjà très bonne pour $T = 2$, presque parfaite pour $T = 4$:



Calcul et estimation de $\tilde{\Delta}$

- D'après ce qui précède :

$$\tilde{\Delta} = \beta_{0k} E \left[\sum_{t=0}^S \frac{a_t(X) \binom{T-t}{S-t} \exp(SX'_T \beta_0)}{C_S(X, \beta_0)} \right].$$

⇒ On l'estime simplement par :

$$\hat{\Delta} = \frac{\hat{\beta}_k}{n} \sum_{i=1}^n \left[\sum_{t=0}^{S_i} \frac{\hat{a}_t(X_i) \binom{T-t}{S_i-t} \exp(S_i X'_{iT} \hat{\beta})}{C_{S_i}(X_i, \hat{\beta})} \right].$$

- On a alors

$$\sqrt{n}(\hat{\Delta} - \tilde{\Delta}) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad (3)$$

pour un $\sigma^2 > 0$ qu'on peut estimer (par $\hat{\sigma}^2$, disons).

L'erreur d'approximation $\tilde{\Delta}$

- Pour faire de l'inférence sur $\Delta \neq \tilde{\Delta}$, on utilise le résultat suivant :

Lemma 1 (Approximation de Δ par $\tilde{\Delta}$)

Supposons (1), alors :

$$|\Delta - \tilde{\Delta}| \leq \bar{b} := \frac{|\beta_{0k}| E(|\lambda_{T+1}(X)| c_0(X))}{2 \times 4^T}.$$

De plus, $|\Delta - \tilde{\Delta}| = \bar{b}$ ssi $\beta_{0k} = 0$ ou une condition très restrictive sur (X, α) est vérifiée (cf. papier).

- On peut alors estimer \bar{b} de manière convergente par :

$$\hat{\bar{b}} = \frac{|\hat{\beta}_k|}{2 \times 4^T} \frac{1}{n} \sum_{i=1}^n |\hat{\lambda}_{T+1}(X_i)| \binom{T}{S_i} \exp(S_i X'_{iT} \hat{\beta}).$$

Construction d'intervalles de confiance sur Δ

- On considère l'intervalle de confiance :

$$\text{IC}_{1-\alpha}^2 = \left[\hat{\Delta} \pm q_\alpha \left(\frac{n^{1/2} \hat{b}}{\hat{\sigma}} \right) \frac{\hat{\sigma}}{n^{1/2}} \right],$$

où $q_\alpha(b) = \text{quantile d'ordre } 1 - \alpha \text{ d'une } |\mathcal{N}(b, 1)|$.

- Intuition : si $\hat{b} = 0$, on n'estimerait pas de biais ($\Delta = \tilde{\Delta}$) et on retrouverait l'IC habituel.
- Sinon $q_\alpha \left(n^{1/2} \hat{b} / \sigma \right) > q_\alpha(0)$: on agrandit l'IC standard pour tenir compte du fait que $\Delta \neq \tilde{\Delta}$.
- On montre que cet IC est valable dès que $|\tilde{\Delta} - \Delta| < \bar{b}$ ou $\beta_{0k} = 0$.

Outline

- 1 Introduction
- 2 Identification
- 3 Estimation de Δ
- 4 Simulations**
- 5 Conclusion

Modèle considéré

- On suppose les X_1, \dots, X_T i.i.d., avec $X_t \in \mathbb{R} \sim \mathcal{U}[-1/2, 1/2]$ et $\beta_0 = 1$.
- On considère $T \in \{2, 3\}$ et $n \in \{250; 500; 1,000\}$.
- On considère trois modèles sur $\alpha|X$, cf. ci-dessous.
- On calcule $IC_{0.95}^1$ issu de la 1ère méthode et $IC_{0.95}^2$.
- Temps de calcul, 1ère méthode (avec un MacBook, Matlab et $n = 1,000$) :
 $T = 2$: 8.1s pour estimer les bornes, 37.7s pour calculer l'IC ;
 $T = 3$: 10.6s pour estimer les bornes, 59.6s pour calculer l'IC.
- Temps de calcul d' $IC_{0.95}^2$: $\simeq 0.02s$!

Modèle 1 : estimation des bornes

- $\alpha = 0$. Δ est identifié ponctuellement pour tout $T \geq 2$,
- Vraie valeur : $2\Lambda(1/2) - 1 \simeq 0.2449$;

T	n	1ère méthode				2ème méthode		
		$\sigma(\hat{\Delta})$	Bias($\hat{\Delta}$)	$\sigma(\hat{\Delta})$	Bias($\hat{\Delta}$)	$\sigma(\hat{\Delta})$	Bias($\hat{\Delta}$)	$E(\hat{b})$
2	250	0.133	0.005	0.138	0.011	0.115	0.0080	0.0136
	500	0.087	0.000	0.090	0.006	0.080	0.0057	0.0119
	1,000	0.066	0.007	0.068	0.013	0.056	0.0048	0.0112
3	250	0.066	-0.011	0.066	-0.010	0.076	0.0027	0.0013
	500	0.044	-0.015	0.044	-0.015	0.055	0.0003	0.0011
	1,000	0.031	-0.013	0.031	-0.013	0.038	0.0007	0.0010

Modèle 1 : comparaison des IC

T	n	$IC_{0.95}^1$		$IC_{0.95}^2$	
		tx de couv.	lgr moyenne	tx de couv.	lgr moyenne
2	250	0.94	0.451	0.96	0.452
	500	0.95	0.318	0.96	0.320
	1,000	0.93	0.225	0.96	0.227
3	250	0.98	0.296	0.95	0.297
	500	0.98	0.208	0.94	0.210
	1,000	0.97	0.146	0.95	0.149

Modèle 2 : estimation des bornes

- Conditionnellement à X , α prend deux valeurs :

$$\alpha = X_T + \eta, \quad P(\eta = -1 | X_1, \dots, X_T) = P(\eta = 1 | X_1, \dots, X_T) = 1/2.$$

- $\Delta \simeq 0.1904$ est partiellement identifié si $T < 4$ et identifiée ponctuellement sinon.
- Bornes : $[0.1826, 0.1953]$ si $T = 2$ et $[0.1895, 0.1906]$ si $T = 3$.

T	n	1ère méthode				2ème méthode		
		$\sigma(\widehat{\Delta})$	Bias($\widehat{\Delta}$)	$\sigma(\widehat{\Delta})$	Bias($\widehat{\Delta}$)	$\sigma(\widehat{\Delta})$	Bias($\widehat{\Delta}$)	$E(\widehat{b})$
2	250	0.116	-0.013	0.122	-0.013	0.098	0.0058	0.0155
	500	0.067	-0.014	0.071	-0.016	0.071	0.0035	0.0131
	1,000	0.048	-0.009	0.050	-0.012	0.048	0.0034	0.0120
3	250	0.065	-0.003	0.066	-0.004	0.066	-0.0013	0.0016
	500	0.047	-0.002	0.047	-0.003	0.047	-0.0001	0.0013
	1,000	0.032	-0.006	0.032	-0.007	0.033	-0.0013	0.0012

Modèle 2 : comparaison des IC

T	n	$IC_{0.95}^1$		$IC_{0.95}^2$	
		tx de couv.	lgr moyenne	tx de couv.	lgr moyenne
2	250	0.93	0.365	0.97	0.395
	500	0.94	0.255	0.96	0.280
	1,000	0.94	0.182	0.97	0.199
3	250	0.97	0.271	0.96	0.260
	500	0.96	0.188	0.95	0.184
	1,000	0.96	0.132	0.95	0.130

Modèle 3 : estimation des bornes

- Conditionnellement à X , α est continue :

$$\alpha = X_T + \eta, \quad \eta | X_1, \dots, X_T \sim \mathcal{N}(0, 1).$$

- $\Delta \simeq 0.1967$ est partiellement identifié pour tout T .
- Bornes pour $T = 2$: $[0.1905, 0.2015]$, pour $T = 3$: $[0.1961, 0.1970]$.

T	n	1ère méthode				2ème méthode		
		$\sigma(\hat{\Delta})$	Bias($\hat{\Delta}$)	$\sigma(\hat{\Delta})$	Bias($\hat{\Delta}$)	$\sigma(\hat{\Delta})$	Bias($\hat{\Delta}$)	$E(\hat{b})$
2	250	0.097	-0.018	0.102	-0.018	0.103	0.0076	0.0155
	500	0.061	-0.020	0.065	-0.022	0.072	0.0063	0.0132
	1,000	0.048	-0.009	0.051	-0.011	0.051	0.0042	0.0120
3	250	0.065	-0.008	0.065	-0.008	0.067	-0.0002	0.0016
	500	0.046	-0.005	0.046	-0.006	0.047	0.0001	0.0013
	1,000	0.035	-0.001	0.035	-0.002	0.033	0.0008	0.0012

Modèle 3 : comparaison des IC

T	n	$IC_{0.95}^1$		$IC_{0.95}^2$	
		tx de couv.	lgr moyenne	tx de couv.	lgr moyenne
2	250	0.95	0.369	0.96	0.404
	500	0.97	0.256	0.96	0.285
	1,000	0.94	0.186	0.96	0.203
3	250	0.96	0.271	0.95	0.261
	500	0.97	0.191	0.95	0.185
	1,000	0.95	0.134	0.95	0.130

Outline

- 1 Introduction
- 2 Identification
- 3 Estimation de Δ
- 4 Simulations
- 5 Conclusion**

Conclusion

- Nous avons obtenu une caractérisation simple de l'AME (et de l'ATE, cf. papier) dans les logit à effets fixes.
- Basée sur cette caractérisation, nous proposons deux méthodes de calcul de Δ .
- La 2ème méthode est très simple. Elle n'est pas optimale en théorie mais fonctionne très bien sur simulations.
- Dans le papier, on étend ces résultats à d'autres paramètres/modèles.
- Nous avons développé, avec Christophe Gaillac (Oxford) :
 1. une commande Stata, `mfelogit`, disponible sur SSC ;
 2. un package R, `MarginalFElogit`, téléchargeable à l'adresse <https://github.com/cgaillac/MarginalFElogit>.N'hésitez pas à les utiliser !
- Lien vers notre document de travail :
<https://arxiv.org/abs/2105.00879>.

Revue de la littérature

Effets marginaux dans les modèles de panel non linéaires avec hétérogénéité inobservée

Modèles paramétriques : Honoré & Tamer (2006), Wooldridge (2019), Aguirregabiria and Carro (2020), Dobronyi et al. (2021)...

Modèles non-paramétriques : Altonji & Matzkin (2005), Hoderlein & White (2012), Chernozhukov, Fernández-Val, Hahn and Newey (2013), Chernozhukov, Fernández-Val, Hoderlein, Holtzmann and Newey (2015)...

Problème des moments

Karlin & Shapley (1953), Krein & Nudelman (1977), Dette & Studden (1997), D'Haultfœuille & Rathelot (2017)... and old results from Chebyshev and Markov !

Inférence prenant en compte le biais

Donoho (1994), Armstrong & Kolesár (2018, 2020, 2021).

Résolution du problème des moments

- Soit $T > 0$ et pour tout $m = (m_0, \dots, m_T) \in \mathbb{R}^{T+1}$,

$$\begin{aligned} \underline{\mathbb{H}}_T(m) &= (m_{i+j-2})_{1 \leq i, j \leq T/2+1}, & \overline{\mathbb{H}}_T(m) &= (m_{i+j-1} - m_{i+j})_{1 \leq i, j \leq T/2} && \text{if } T \text{ even} \\ \underline{\mathbb{H}}_T(m) &= (m_{i+j-1})_{1 \leq i, j \leq (T+1)/2}, & \overline{\mathbb{H}}_T(m) &= (m_{i+j-2} - m_{i+j-1})_{1 \leq i, j \leq (T+1)/2} && \text{if } T \text{ odd.} \end{aligned}$$

- Soit alors $\underline{H}_T(c) = \det(\underline{\mathbb{H}}_T(c))$ et $\overline{H}_T(c) = \det(\overline{\mathbb{H}}_T(c))$.

Proposition 1 (déterminants de Hankel)

$m \in \mathcal{M}_T$ ssi $\underline{H}_T(m) \times \overline{H}_T(m) \geq 0$. De plus,

1. Si $\underline{H}_T(m) \times \overline{H}_T(m) > 0$, $\underline{q}_T(m) < \overline{q}_T(m)$. De plus, $q \mapsto \underline{H}_{T+1}(m, q)$ est strictement \uparrow , linéaire et $\underline{H}_{T+1}(m, \underline{q}_T(m)) = 0$. Idem pour $\overline{q}_T(m)$.
2. Si $\underline{H}_T(m) \times \overline{H}_T(m) = 0$, on a $\underline{q}_T(m) = \overline{q}_T(m)$. De plus, en notant $T' = \min\{t \leq T : \underline{H}_t(m) \times \overline{H}_t(m) = 0\}$, $\underline{q}_T(m) = \overline{q}_T(m)$ vérifie

$$\begin{aligned} \underline{H}_{T'}(m_{T-T'+1}, \dots, m_T, \underline{q}_T(m)) &= 0 && \text{if } \underline{H}_{T'}(m) = 0, \\ \overline{H}_{T'}(m_{T-T'+1}, \dots, m_T, \underline{q}_T(m)) &= 0 && \text{if } \overline{H}_{T'}(m) = 0. \end{aligned}$$

Interprétation géométrique

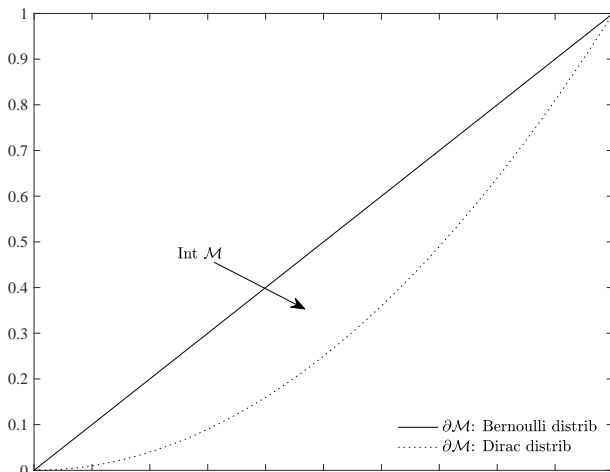


Figure – Illustration de l'espace des moments pour $T = 2$