Monotonic Effects of Characteristics on Returns

Jared Fisher
with Carlos Carvalho, David Puelz

The University of Texas
McCombs School of Business
Department of Information, Risk and Operations Management

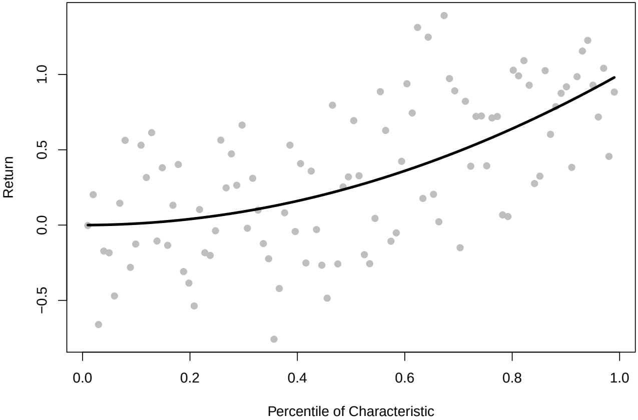
June 26, 2018

Empirical Asset Pricing

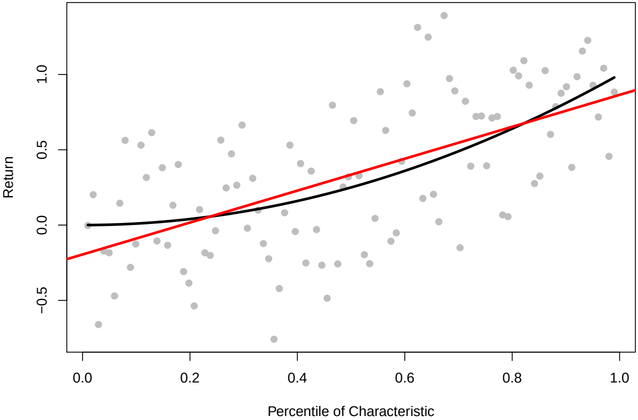
$$\mathbb{E}(R_t|\boldsymbol{x}_{t-1}) = f(\boldsymbol{x}_{t-1})$$

- What should be included in $\boldsymbol{x}_{t-1}$, given f?
- We also want to ask, what is f?
- What if f changes over time?

Example Data



Linear Regression



Linear Regression, e.g. Fama and French (1992)

Table III

Average Slopes (t-Statistics) from Month-by-Month Regressions of Stock Returns on β, Size, Book-to-Market Equity, Leverage, and E/P: July 1963 to December 1990

Stocks are assigned the post-ranking β of the size-β portfolio they are in at the end of June of year t (Table I). BE is the book value of common equity plus balance-sheet deferred taxes, A is total book assets, and E is earnings (income before extraordinary items, plus income-statement deferred taxes, minus preferred dividends). BE, A, and E are for each firm's latest fiscal year ending in calendar year $t-1$. The accounting ratios are measured using market equity ME in December of year $t-1$. Firm size ln(ME) is measured in June of year t. In the regressions, these values of the explanatory variables for individual stocks are matched with CRSP returns for the months from July of year t to June of year $t+1$. The gap between the accounting data and the returns ensures that the accounting data are available prior to the returns. If earnings are positive, E(+)/P is the ratio of total earnings to market equity and E/P dummy is 0. If earnings are negative, E(+)/P is 0 and E/P dummy is 1.

The average slope is the time-series average of the monthly regression slopes for July 1963 to December 1990, and the t-statistic is the average slope divided by its time-series standard error.

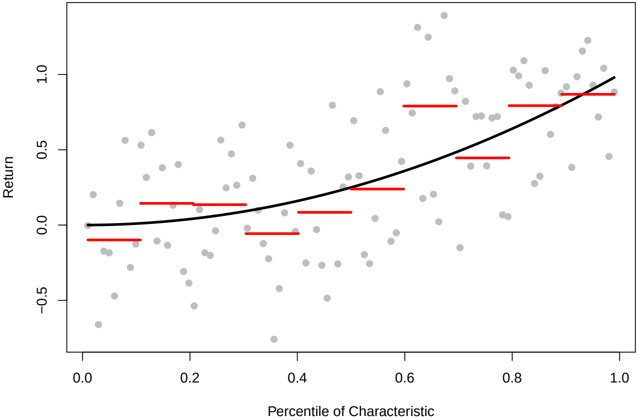
On average, there are 2267 stocks in the monthly regressions. To avoid giving extreme observations heavy weight in the regressions, the smallest and largest 0.5% of the observations on E(+)/P, BE/ME, A/ME, and A/BE are set equal to the next largest or smallest values of the ratios (the 0.005 and 0.995 fractiles). This has no effect on inferences.

β	ln(ME)	ln(BE/ME)	ln(A/ME)	ln(A/BE)	E/P Dummy	E(+)/P
0.15 (0.46)						
	−0.15 (−2.58)					
−0.37 (−1.21)	−0.17 (−3.41)					
		0.50 (5.71)				
			0.50 (5.69)	−0.57 (−5.34)		
					0.57 (2.28)	4.72 (4.57)
	−0.11 (−1.99)	0.35 (4.44)				
	−0.11 (−2.06)		0.35 (4.32)	−0.50 (−4.56)		
	−0.16 (−3.06)				0.06 (0.38)	2.99 (3.04)
	−0.13 (−2.47)	0.33 (4.46)			−0.14 (−0.90)	0.87 (1.23)
	−0.13 (−2.47)		0.32 (4.28)	−0.46 (−4.45)	−0.08 (−0.56)	1.15 (1.57)

Linear Regression

Freyberger, Neuhierl, and Weber (2017): "no *a priori* reason exists why the conditional mean function should be linear."

Nonlinear Option: Portfolio Sorts



Portfolio Sorts, e.g. Jegadeesh and Titman (2001)

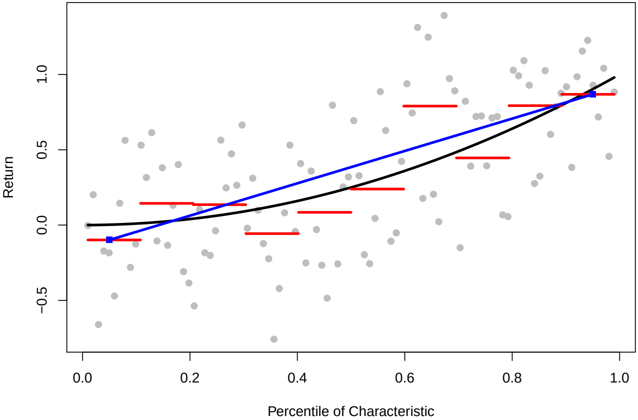
Table I

Momentum Portfolio Returns

This table reports the monthly returns for momentum portfolios formed based on past six-month returns and held for six months. P1 is the equal-weighted portfolio of 10 percent of the stocks with the highest returns over the previous six months, P2 is the equal-weighted portfolio of the 10 percent of the stocks with the next highest returns, and so on. The "All stocks" sample includes all stocks traded on the NYSE, AMEX, or Nasdaq excluding stocks priced less than $5 at the beginning of the holding period and stocks in the smallest market cap decile (NYSE size decile cutoff). The "Small Cap" and "Large Cap" subsamples comprise stocks in the "All Stocks" sample that are smaller and larger than the median market cap NYSE stock respectively. "EWI" is the returns on the equal-weighted index of stocks in each sample.

	All Stocks			Small Cap			Large Cap		
	1965–1998	1965–1989	1990–1998	1965–1998	1965–1989	1990–1998	1965–1998	1965–1989	1990–1998
P1 (Past winners)	1.65	1.63	1.69	1.70	1.69	1.73	1.56	1.52	1.66
P2	1.39	1.41	1.32	1.45	1.50	1.33	1.25	1.24	1.27
P3	1.28	1.30	1.21	1.37	1.42	1.23	1.12	1.10	1.19
P4	1.19	1.21	1.13	1.26	1.34	1.05	1.10	1.07	1.20
P5	1.17	1.18	1.12	1.26	1.33	1.06	1.05	1.00	1.19
P6	1.13	1.15	1.09	1.19	1.26	1.01	1.09	1.05	1.20
P7	1.11	1.12	1.09	1.14	1.20	0.99	1.09	1.04	1.23
P8	1.05	1.05	1.03	1.09	1.17	0.89	1.04	1.00	1.17
P9	0.90	0.94	0.77	0.84	0.95	0.54	1.00	0.96	1.09
P10 (Past losers)	0.42	0.46	0.30	0.28	0.35	0.08	0.70	0.68	0.78
P1–P10	1.23	1.17	1.39	1.42	1.34	1.65	0.86	0.85	0.88
t statistic	6.46	4.96	4.71	7.41	5.60	5.74	4.34	3.55	2.59
EWI	1.09	1.10	1.04	1.13	1.19	0.98	1.03	1.00	1.12

Portfolio Sorts



Portfolio Sorts

Fama and French (2008): "sorts are clumsy for examining the functional form of the relation between average returns and an anomaly variable."

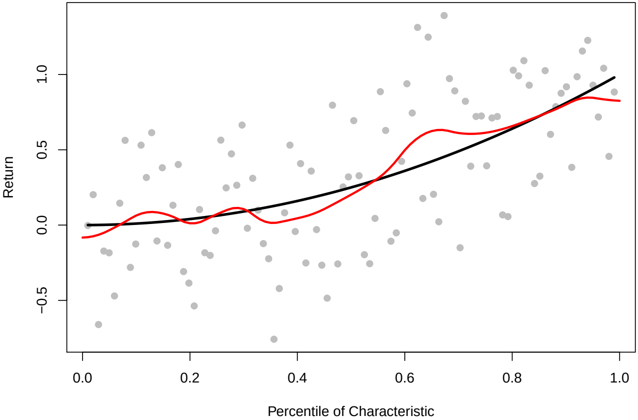
- Function is assumed constant within deciles
- No information shared across deciles

Portfolio Sorts

Fama and French (2008): "sorts are clumsy for examining the functional form of the relation between average returns and an anomaly variable."

- Function is assumed constant within deciles
- No information shared across deciles
- (There are ways to fit nonlinear functions that are smooth)

Splines



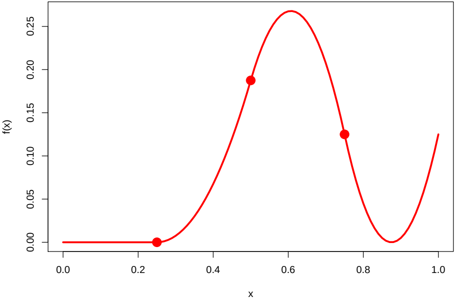
Return
Percentile of Characteristic
−0.5
0.0
0.5
1.0
0.0
0.2
0.4
0.6
0.8
1.0

Splines

For a standard spline with m knots, $\tilde{x}_1, ..., \tilde{x}_m$, $x \in [0,1]$,

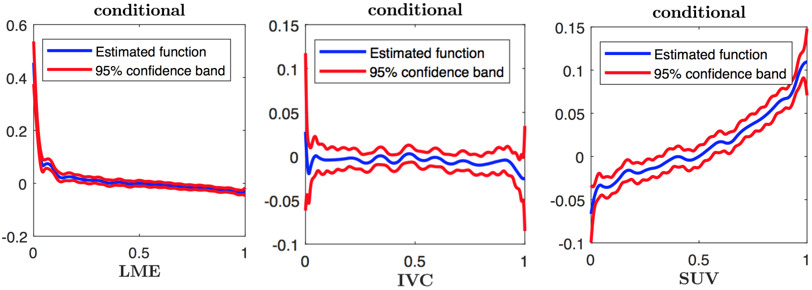
$$f(x) = \alpha + \beta_1 x + \beta_2 x^2 + \beta_3 (x - \tilde{x}_1)_+^2 + ... + \beta_{m+2}(x - \tilde{x}_m)_+^2$$

Fits a smooth curve with a different quadratic coefficient after each knot. For example, knots at 0.25, 0.5, 0.75,

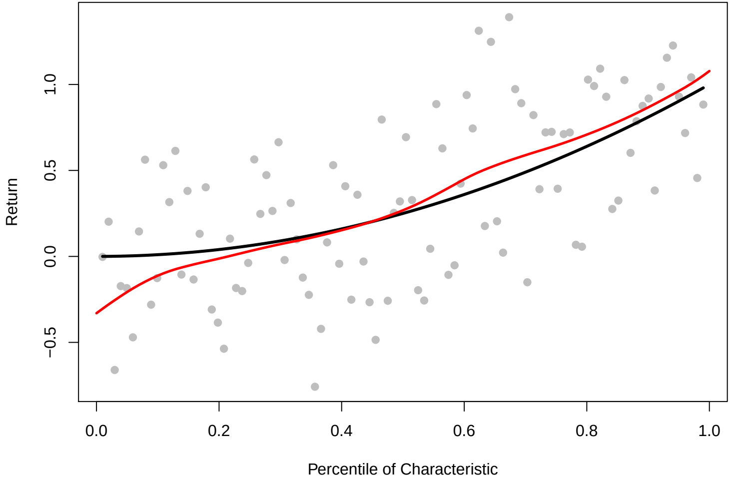


Splines - Freyberger, Neuhierl, and Weber (2017)

- Characteristics/covariates are rank-transformed to empirical percentiles, in (0,1)
- Additive model of quadratic splines, but include separate intercepts for each covariate
- Fit with Adaptive Group LASSO, which shrinks and selects all of a characteristic's spline coefficients as a group



Monotonic Quadratic Spline



Our Contribution

If we are serious about understanding the functional form of these marginal relationships, then we should have

1. Additive splines: flexible and can separate to marginal effects
2. Monotonicity: complement the flexibility of the splines with a priori known structure
3. A single intercept: identifiable and intuitive
4. Time-dynamics modeled, not just a rolling window
5. Separation between the shrinkage of coefficients and selection of characteristics

1 - Additive Model

$$\mathbb{E}(r_{it}|\boldsymbol{x}_{i,t-1}) = \alpha_t + \sum_{k=1}^{K} f_{kt}(x_{k,i,t-1})$$

1 - Additive Model

$$\mathbb{E}(r_{it}|\boldsymbol{x}_{i,t-1}) = \alpha_t + \sum_{k=1}^{K} f_{kt}(x_{k,i,t-1})$$

- $x_{k,i,t-1} \in (0,1)$ is the empirical percentile of characteristic k for firm i at time $t-1$, ranked over all firms

1 - Additive Model

$$\mathbb{E}(r_{it}|\boldsymbol{x}_{i,t-1}) = \alpha_t + \sum_{k=1}^{K} f_{kt}(x_{k,i,t-1})$$

- $x_{k,i,t-1} \in (0,1)$ is the empirical percentile of characteristic k for firm i at time $t-1$, ranked over all firms
- Note that there are no interactions built into the model, as the intention is to see the marginal effect

2 - Monotonicity: Shively, Sager and Walker (2009)

A standard spline with m knots, $\tilde{x}_1, ..., \tilde{x}_m$,

$$f(x) = \beta_1 x + \beta_2 x^2 + \beta_3 (x - \tilde{x}_1)_+^2 + ... + \beta_{m+2}(x - \tilde{x}_m)_+^2$$

is forced to be (WLOG) nondecreasing if all first derivatives are nonnegative. These simple derivatives, with known knots, yield $m + 2$ linear constraints needed for monotonicity, such that

$$\boldsymbol{L\beta} \geq 0.$$

So, we set $\boldsymbol{\gamma} = \boldsymbol{L\beta}$ and we use a modified version of their shrinkage prior:

$$\begin{aligned}(\gamma_j | I_j = 0) &\sim \delta_0 \\ (\gamma_j | I_j = 1) &\sim N_+(0, c\sigma^2) \\ I_j &\sim Bernoulli(0.2)\end{aligned}$$

3 - Intercept adjustment

Recall our additive model, with spline basis $\boldsymbol{X}_{i,k,t-1}$ and a single intercept

$$\mathbb{E}(r_{it}|\boldsymbol{x}_{i,t-1}) = \alpha_t + \sum_{k=1}^{K} \boldsymbol{X}_{i,k,t-1}\boldsymbol{\beta}_{kt}$$

$\Rightarrow$ α_t is the expected return for a firm with the minimum value for all characteristics, i.e. $\boldsymbol{X}_{i,k,t-1} = 0, \forall k$.
Problems:

1. Computationally challenging due to few and volatile data points
2. Intuitively unfavorable as a baseline
3. Cannot see the lower tail effects change over time

3 - Intercept adjustment

Proposal: let the intercept be the expected return for a firm that has the median value for all characteristics

- Requires transforming the splines such that they equal 0 at the median $x = 0.5$ and not $x = 0$
- This then requires carefully expand spline basis and the monotonicity constraint matrix L

4 - Modeling Time-dynamics: McCarthy and Jensen (2016)

- Power-weighted likelihoods let information decay over time
- To estimate parameters at time τ, let $\delta_t = 0.99^{\tau - t}$, such that $\delta_1 \leq \delta_2 \leq ... \leq \delta_\tau = 1$, the likelihood at time $\tau \in \{1, ..., T\}$ is

$$p(\boldsymbol{r}_1, ..., \boldsymbol{r}_\tau | \Theta_\tau) = \prod_{t=1}^{\tau} p(\boldsymbol{r}_t | \Theta_\tau)^{\delta_t}.$$

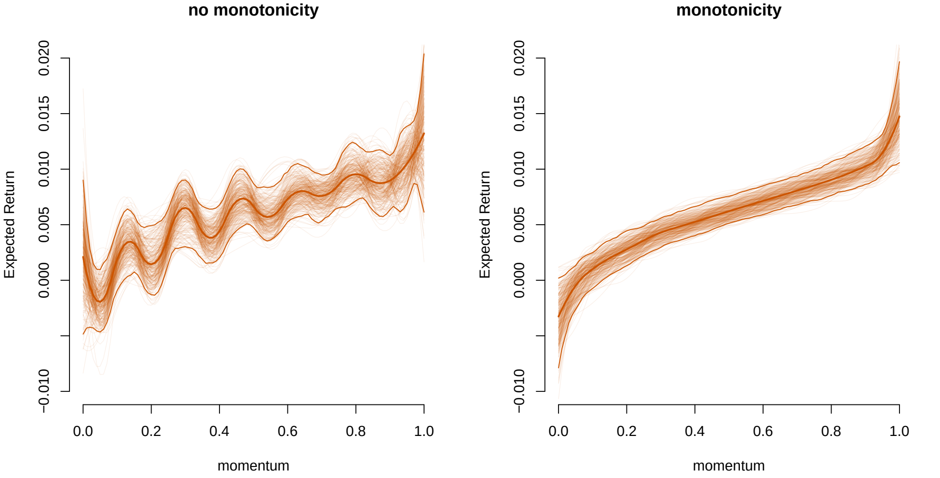
Data

Freyberger, Neuhierl, and Weber (2017)'s dataset:

- CRSP monthly stock returns for most US traded firms
- 36 characteristics from Compustat and CRSP, including size, momentum, leverage, etc.
- July 1962 - June 2014

Presence and direction of monotonicity is determined by important papers in the literature

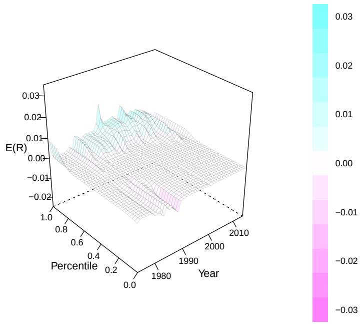
Estimated functions at January 1978



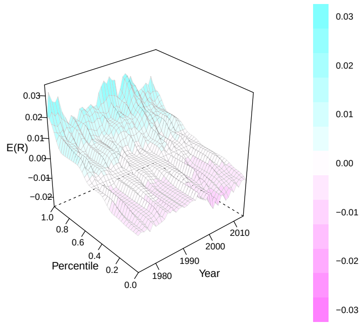
*unconditional

How does the function vary over time?

Momentum (r_{12-2})

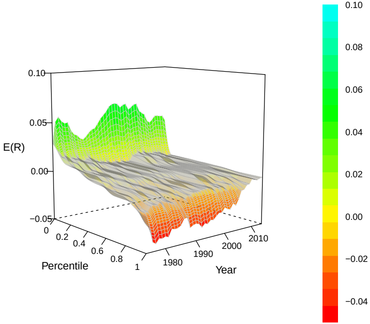


Standard Unexplained Volume

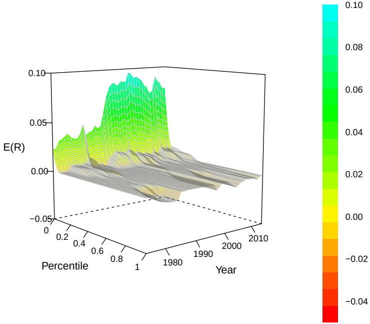


How does the function vary over time?

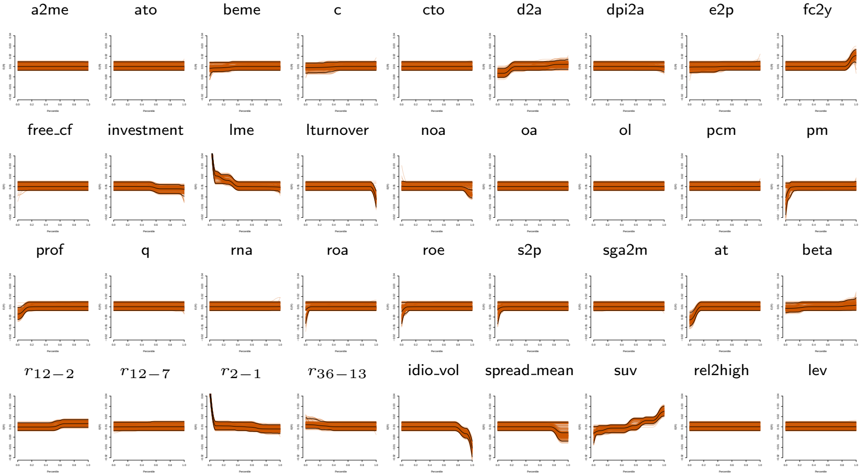
Short-term Reversal (r_{2-1})



Size



Keep Only the “Significant?” - January 2014



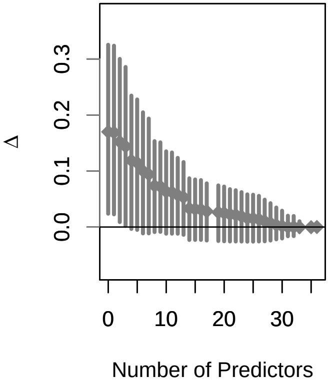
5 - Utility-based Variable Selection: Puelz, Hahn, and Carvalho (2017,2018)

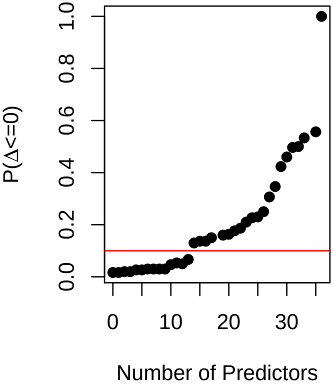
1. Specify utility function: model fit + complexity penalty
2. Optimize expected utility:
   a. Integrate over $(\tilde{R}_t, \Theta_t)$

   $$\mathcal{L}_{\lambda_t}(\mathbf{A}_t) = \left\| \mathbb{X}_{t-1}\mathbf{A}_t - \mathbb{X}_{t-1}\overline{\mathbf{B}}_t \right\|_2^2 + \Phi(\lambda_t, \mathbf{A}_t)$$

   b. Optimize for a given λ_t
3. Compare optimal sparse models in light of uncertainty. We care about the difference in utility between optimal sparse models and the dense model

Posterior Summarization - January 2014





Selected: beme, c, d2a, investment, lme, prof, s2p, sga2m, r_{12-2}, r_{2-1}, r_{36-13}, idio_vol, suv, lev

Conclusion

- We present a model that fits expected excess returns as flexible functions of firm characteristics
- These functions can include a priori knowledge of monotonicity
- These function are dynamic and adapt over time
- We will continue to develop the variable selection process

Appendix: Model Summary

$$
\begin{aligned}
\boldsymbol{r}_t|\cdot &\sim N\left(\alpha_t \mathbf{1}_{n_t} + \sum_{k=1}^{K} f_{kt}(\boldsymbol{x}_{k,t-1}),\ \sigma_t^2 I_n\right)^{\delta_t} \\
f_{kt}(\boldsymbol{x}_{k,t-1}) &= X_{k,t-1}\boldsymbol{\beta}_{kt} = X_{k,t-1}L^{-1}L\boldsymbol{\beta}_{kt} = W_{kt}\boldsymbol{\gamma}_{kt} \\
\alpha_t &\sim N(0, 10^{-2}) \\
\sigma_t^2 &\sim U(0, 10^3) \\
(\gamma_{jkt}|I_{jkt} = 1, \sigma_t^2) &\sim N_+(0, c_k\sigma_t^2) \\
(\gamma_{jkt}|I_{jkt} = 0) &= 0 \\
I_{jkt} &\sim Bn(p_{jk} = 0.2).
\end{aligned}
$$