Online Supplement to
"Structural Estimation of Behavioral Heterogeneity"

Zhentao Shi and Huanhuan Zheng

Chinese University of Hong Kong and National University of Singapore

February 10, 2018

Due to space limitation, we prepare this Online Supplement for robustness check, additional empirical results, some implementation details, an extension of the model, and two more examples of the heterogeneous agent model to which the technique of thin-set identification is applicable.

S1 Robustness Check: ELXM

Empirical likelihood (Qin and Lawless, 1994; Kitamura, 1997) is an alternative to GMM. It is natural to design empirical likelihood with extended moments (ELXM) as a counterpart of extended method of moments (XMM). To check the robustness of the estimation results across different methods, we estimate the model with ELXM in this section. We first describe how to carry out ELXM.

If the observations are i.i.d., empirical likelihood (EL) is formulated as a constrained optimization problem

$$\max_{\theta\in\Theta,(\pi_t\in[0,1])_{t=1}^T} \sum_{t=1}^T \log \pi_t, \quad \text{subject to} \quad \sum_{t=1}^T \pi_t = 1 \text{ and } \sum_{t=1}^T \pi_t \mathbf{g}_t(\theta) = 0,$$

where π_t is the probability assigned to the t-th observation. EL is known to be asymptotically equivalent to GMM at the first order.

In time series, however, the blockwise EL (Kitamura, 1997) takes a distinctive scheme to account for the temporal dependence. We propose (blockwise) ELXM for time series. Let B_T be the block size and $S = \lfloor T/B_T \rfloor$ be the number of blocks. The blockwise moment function can be written as

$$g_{js}^{(B_T)}(\theta) = \frac{1}{B_T} \sum_{t=s(B_T-1)+1}^{sB_T} g_{jt}(\theta), \text{ for } s = 1, \ldots, S; j = 1, \ldots, 8$$

where the blockwise summation deals with time dependence. The primal problem of ELXM is formulated as

$$\max_{\theta \in \Theta, (\pi_s \in [0,1])_{s=1}^S} \sum_{s=1}^S \log \pi_s \text{ subject to } \sum_{s=1}^S \pi_s = 1 \text{ and } \sum_{s=1}^S \pi_s \mathbf{g}_s^{(B_T)}(\theta) = 0, \quad \text{(S1)}$$

where $\mathbf{g}_s^{(B_T)}(\theta)$ is the blockwise counterpart of $\mathbf{g}_t(\theta)$, and π_s is the probability assigned to the s-th block. We denote the maximizer of θ in (S1) as $\widehat{\theta}_{\text{ELXM}}$. ELXM is the adaption of XMM into the EL framework. Stable results between ELXM and XMM would reinforce the robustness to the numerical optimization procedure and the tuning parameters for time dependence.

As an extension of XMM, Gagliardini et al. (2011, pp.2109–2010) have discussed the asymptotic distribution of XMM's EL cousin. If $B_T \to \infty$, $B_T/T^{1/2} \to 0$ as $T \to \infty$ (Kitamura, 1997, Theorem 1(vii), p.2090), the asymptotic distribution of $\widehat{\theta}_{\text{ELXM}}$ is equivalent to that of $\widehat{\theta}_{\text{XMM}}$. First-order asymptotic equivalence further indicates that the likelihood ratio statistic

$$LR = 2\left(S \log\left(\frac{1}{S}\right) - \sum_{s=1}^S \log \widehat{\pi}_s\right) \xrightarrow{\text{d}} \chi^2(4),$$

where $(\widehat{\pi}_s)_{s=1}^S$ is the implied probability—the maximizer of the $(\pi_s)_{s=1}^S$ part in the primal problem (S1). Therefore, ELXM estimator is asymptotic equivalent to XMM, and the likelihood ratio test follows the same asymptotic distribution as that of the J test.

The numerical implementation of ELXM is similar to the standard blockwise EL. We carry out the numerical optimization in two steps: (i) solve π in the inner step given a trial value θ,

Table S1: Estimation Results of ELXM for the Full Model

	Period 1		Period 2		Period 3	
	est.	95% CI	est.	95% CI	est.	95% CI
σ_μ	0.013	(0.010, 0.017)	0.007	(0.006, 0.008)	0.028	(0.028, 0.029)
η	0.106	(0.087, 0.126)	0.167	(0.125, 0.210)	0.218	(0.126, 0.310)
τ	0.597	(0.430, 0.763)	0.706	(0.422, 0.989)	0.857	(0.415, 1.300)
α	1.551	(0.642, 2.461)	1.863	(0.946, 2.779)	3.240	(0.808, 5.673)
LR-stat.	4.205		5.121		8.684	
p-value	(0.379)		(0.275)		(0.069)	

Note: Similar to Table 1, the likelihood ratio statistics (LR-stat.) of the over-identification test follows $\chi^2(4)$ asymptotic distribution under the null.

and (ii) solve θ in the outer step. We optimize the convex primal problem in the inner loop, while the outer step is a standard low-dimensional nonlinear optimization. We set B_T equal to the number of lags for the long-run variance calculation in XMM in Section 3.4 of the main text.

The ELXM estimation of the full model is reported in Table S1. The point estimates and the confidence intervals are close to those of XMM. The point estimates also yield very similar predicted moments as XMM in Table 2 and the switching between fundamentalists and chartists exhibits similar patterns with those in Figures 1, 2, and 3, which we do not repeat here. Nevertheless, the LR test statistics of Period 3 is 8.68, with a p-value of 0.07. The over-identification test rejects the null hypothesis at 10% significance level. The evidence of marginal rejection echoes the big J-statistic for Period 3 in Table 1 of the main text. It indicates that we must be cautious when applying our model to a long time span with high volatility and potential structural changes.

For further comparison, we run the standard blockwise EL to estimate the model with unconditional moments, as we did for GMM. The results are displayed in Table S2. Again, we observe the pattern of smaller η and wider confidence intervals, which echoes that in Table 3. We also try ELXM for the solo-strategy models, in which we encounter the numerical problem of infeasible constraints in all three periods. The infeasibility problem is well understood in the literature of EL as strong evidence of model misspecification (Chen et al., 2008). Severe model misspecification is manifest in the very small p-values in Table 4. The evidence from the ELXM and EL estimation suggests robustness of the empirical results in the full model and the model with the unconditional moments, as well as strong rejection of the solo-strategy models.

Table S2: Estimation Results of EL for the Unconditional Moment Model

	Period 1		Period 2		Period 3	
	est.	95% CI	est.	95% CI	est.	95% CI
σ_μ	0.014	(0.014, 0.015)	0.007	(0.007, 0.007)	0.029	(0.029, 0.030)
η	0.116	(0.007, 0.225)	0.126	(-0.218, 0.470)	0.060	(-0.004, 0.124)
τ	0.678	(0.439, 0.917)	0.625	(0.116, 1.133)	0.750	(0.421, 1.080)
α	2.644	(0.879, 4.408)	1.751	(-0.135, 3.636)	3.672	(2.200, 5.145)
LR-stat.	0.037		1.103		0.234	
p-value	(0.848)		(0.294)		(0.628)	

Note: Similar to Table 3, the likelihood ratio statistic of the over-identification test follows $\chi^2(1)$ asymptotic distribution under the null.

S2 Additional Empirical Results

In the main text, valid inference relies on several assumptions in the structural model. This section presents additional empirical results to verify some assumptions.

Local identification. In the main text we have assumed local identification, following Gagliardini et al. (2011) and Antoine and Renault (2012). Here we provide statistical evidence of local identification. Local identification is equivalent to a full-rank Jacobian matrix. We use Kleibergen and Paap (2006)'s reduced-rank test (KP test) to check the rank of the empirical Jacobian matrix $\widehat{H}_{\text{unc}}(\theta) = \frac{\partial}{\partial\theta'}\overline{\mathbf{g}}_{\text{unc}}(\theta)$, where $\overline{\mathbf{g}}_{\text{unc}} = \left(\overline{\mathbf{g}}_j\right)_{j\in\{1,5,\ldots,8\}}$ is the vector of the 5 unconditional sample moments. The data support a full rank $\widehat{H}_{\text{unc}}(\theta)$ if we can reject the null hypothesis that its rank is 3, 2, or 1. We evaluate the rank of $\widehat{H}_{\text{unc}}(\theta)$ at either $\widehat{\theta}_{\text{XMM}}$ or $\widehat{\theta}_{\text{GMM}}$. Table S3 reports the KP test statistics under the null of rank 3. We have also conducted the same test under the null that the rank of $\widehat{H}_{\text{unc}}(\theta)$ is 2 or 1, respectively, and the rejection is overwhelming in all cases. The KP test provides evidence of non-trivial local information from the unconditional moments.

Table S3: KP Test Statistic and p-value

	Period 1	Period 2	Period 3
XMM	15.850 (0.000)	5.826 (0.054)	7.319 (0.026)
GMM	19.141 (0.000)	7.251 (0.027)	9.563 (0.008)

Note: The null hypothesis is that the rank of $\widehat{H}_{\text{unc}}(\theta)$ is 3. The p-value in the parenthesis is calculated according to the asymptotic distribution $\chi^2(2)$.

Table S4: Correlation Coefficients of the Real and Predicted returns

		Real	XMM full	GMM	XMM fund.-only
Period 1	XMM full model	0.047			
	GMM	0.048	0.975		
	XMM fundamentalist-only	0.075	0.613	0.565	
	XMM chartist-only	0.165	0.279	0.281	-0.197
Period 2	XMM full model	0.117			
	GMM	0.127	0.949		
	XMM fundamentalist-only	0.092	0.582	0.484	
	XMM chartist-only	0.081	0.201	0.313	-0.374
Period 3	XMM full model	0.112			
	GMM	0.130	0.821		
	XMM fundamentalist-only	0.014	0.410	0.077	
	XMM chartist-only	0.151	0.173	0.473	-0.538

Note: the time series here are the same as those in Table 2. The the entries are pairwise correlation coefficient.

Unit root test for $\boldsymbol{\mu}^T$. The 1%, 5%, and 10% critical value for the standard Dicky-Fuller test are -2.58, -1.95, -1.62 respectively. This is a one-sided test that rejects the null of unit root behavior if the test statistic is smaller than the critical value. We run the Dicky-Fuller test, and obtain the test statistics 0.0702, 0.9672, and -0.4081 for Period 1, 2 and 3, respectively. These statistics are in favor of the null hypothesis of the unit root. Formally, they do not reject the null of unit root at 10% significance level, since none of the statistics are smaller than -1.62. What is more, the positive statistics in period 1 and 2, which are associated with autoregressive coefficient estimates of 1.0004 and 1.0024, respectively, may indicate possibly very weak explosive behavior.

Correlation. Table S4 reports the pairwise correlation coefficients of the real return time series ($\boldsymbol{R}_t^{\mathrm{r}}$) and the predicted ($R_t(\theta)$) evaluated at the various estimates. The entries of the first column of Table S4 are small, indicating weak correlation between the real return and the predicted ones. This is not surprising since we fit the moments of the marginal distribution of the returns, rather than the temporal co-movements, to estimate the parameters.

Frequency of the event G_1. The analysis of the thin-set identification starts from the event $G_1 = \{\Delta_{t-1} = 0\}$, and the convergence rate of the local moments depends on how often Δ_{t-1} fluctuates around 0. Figure S1 plots the series $(\Delta_t)_{t=1}^T$ in all the time periods. We observe the curve vacillates around 0 repeatedly, so that G_1 is not a rare event.

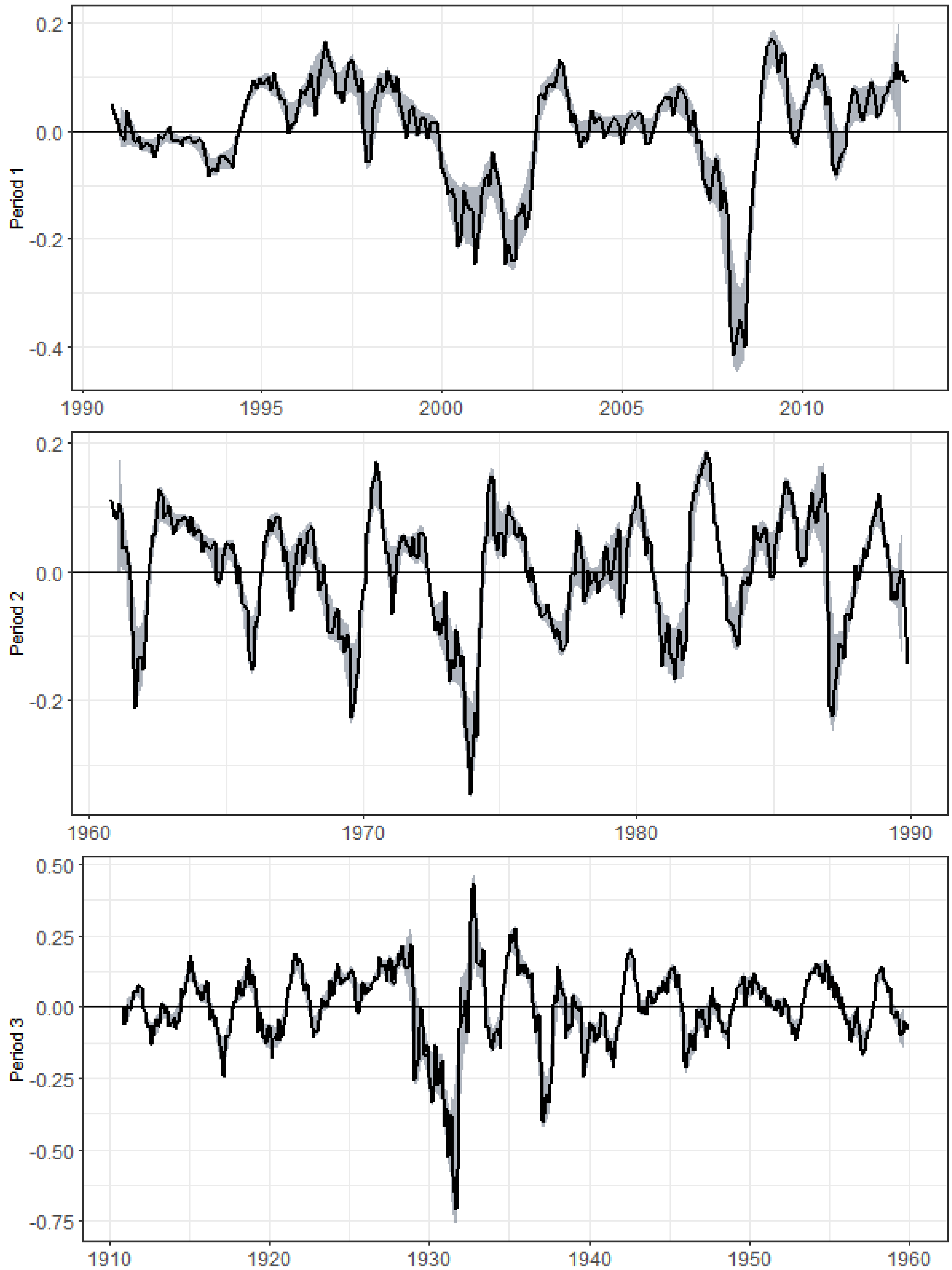
Table S5: Two-regime Markov Switching Model

		Period 1		Period 2		Period 3	
		est.	s.e.	est.	s.e.	est.	s.e.
Regime 1: Boom	Intercept	0.183	0.017	0.148	0.008	-0.003	0.009
	Slope	-0.481	0.180	0.679	0.088	1.864	0.033
Regime 2: Bust	Intercept	-0.276	0.010	-0.198	0.011	-0.132	0.013
	Slope	-0.824	0.101	0.348	0.123	0.446	0.038
Transition	Boom→Bust	0.018		0.020		0.025	
Probability	Bust→Boom	0.019		0.015		0.021	

Markov Switching. We conduct a simple Markov switching model in which we allow two regimes for the intercept and the slope coefficient in the regression $R_t^{\mathrm{r}} = \text{intercept} + \text{slope} \times \mu_t + \text{error term}$. We refer to the regime with greater intercept and slope coefficient as the *boom regime* and the other as the *bust regime*. The boom regimes for sample period 1, 2 and 3 are shaded in yellow color in the upper, middle and bottom panel of Figure S2, respectively.

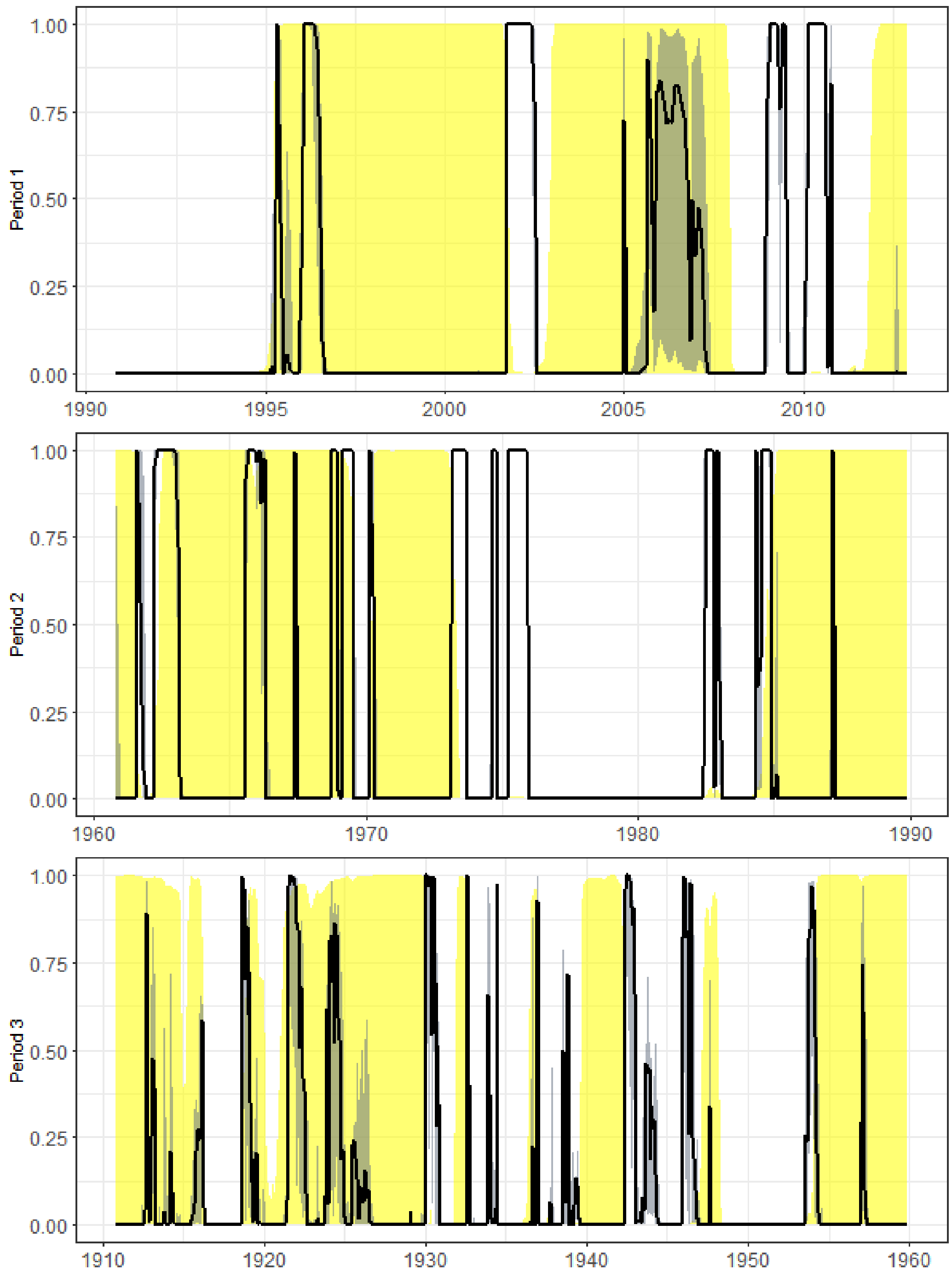
The boom (bust) regimes correspond to the scenarios when the market price is rising (falling). The probability for the market to transit from a boom to a bust ranges from 1.8% to 2.5%, which implies that on average it takes 40 to 56 months for the price to reverse its trend. Similarly, the probability for the market to transit from a bust to a boom is very low.

There are considerable overlap between the boom regimes identified by the Markov switching model and the chartists-dominated regime uncovered from the structural model. It suggests that our model based on the dynamic transition in the market fraction of chartists reasonably captures the price movement.



Note: The gray shaded region is the 90% pointwise confidence interval constructed by the time series kernel smoothing method (Fan and Yao, 2003, p.218). We use the Bartlett kernel with the same bandwidth as in the main text.

Figure S1: $(\Delta_t)_{t=1}^{T}$ in All the Three Periods



Note: The gray shaded region is the 90% pointwise confidence interval constructed by parametric bootstrap. The yellow shaded region is the probability of the boom regime estimated from the two-regime Markov switching model.

Figure S2: Fraction of Chartists and Markov Switching

Confidence interval of $m_t\left(\widehat{\theta}_{\mathrm{XMM}}\right)$. The fraction of the chartist, $m_t(\theta_0)$, is a nonlinear function of $\mathbf{p}^T$ and $\boldsymbol{\mu}^T$. In principle we can construct the pointwise confidence interval by the delta method based on the asymptotic distribution of $\widehat{\theta}_{\mathrm{XMM}}$. However, it is difficult to interpret when the two-sided symmetric confidence interval goes beyond $[0,1]$, which occurred in our experiment.

To avoid such difficulty, we can use the parametric bootstrap if we are willing to impose the normality assumption $\varepsilon_t^{\mu} \sim$ i.i.d.$N(0,1)$. Let $\widehat{\theta}_{\mathrm{XMM}}^{*(b)}$ be a bootstrap estimator where the superscript "(b)" indexes the instance of bootstrap replication, and $m_t^{*(b)} = m_t\left(\widehat{\theta}_{\mathrm{XMM}}^{*(b)}\right)$ is the plug-in bootstrap estimator of the fraction. The parametric bootstrap is implemented as follows. We simulate a sequence $\boldsymbol{\mu}^{T*(b)} = \left(\mu_t^{*(b)}\right)_{t=1}^{T}$ where $\mu_t^{*(b)} = \mu_{t-1} + \widehat{\sigma}_{\mu,\mathrm{XMM}}\varepsilon_t^{\mu*(b)}$ with $\varepsilon_t^{\mu*(b)} \sim$ i.i.d.$N(0,1)$. Given the data $\left(\boldsymbol{\mu}^{T*(b)}, \mathbf{p}^T\right)$, we obtain the bootstrap estimator $\widehat{\theta}_{\mathrm{XMM}}^{*(b)}$. Here we only bootstrap $\boldsymbol{\mu}^{T*(b)}$ since the function $R_t(\theta)$ only depends on the $\left(p_{t-1}, p_{t-1}^c, \mu_t, \mu_{t-1}\right)$ but not p_t. After having $\widehat{\theta}_{\mathrm{XMM}}^{*(b)}$, we plug it into $m_t(\theta)$ and get $m_t^{*(b)}$. We repeat the bootstrap for 199 times, and compute the 5% and 95% sample quantiles of $\left(m_t^{*(b)}\right)_{b=1}^{200}$ for each t as the lower and upper bounds of the 90% two-sided pointwise confidence interval for $m_t(\theta_0)$.

The estimated pointwise bootstrap confidence interval is shown as the gray shaded region in Figure S2. The confidence interval is very narrow most of the time, in particular when $m_t\left(\widehat{\theta}_{\mathrm{XMM}}\right)$ is close to 0. Interestingly, along with the swings of the fraction of the chartists before the 2008 financial crisis, the uncertainty is manifest by the relatively wide confidence intervals.

S3 Implementation

S3.1 Gordon Growth Model

The original Gordon growth model is defined as $\tilde{\mu} = d_t(1+\kappa)/(\beta-\kappa)$, where d_t is the dividend at period t, β is the discount rate and κ is the average growth rate of dividends. Fama and French (2002) suggest that the Gordon growth model implies $\beta = \bar{y} + \kappa$, where $\bar{y}$ is the average dividend yield. We replace β by $\bar{y}+\kappa$ and obtain $\mu_t = d_t(1+\kappa)/\bar{y}$.

S3.2 Chartist-Only Model

Unlike the fundamentalist-only model, the chartist-only model is not a sub-model of the benchmark model, since η cannot be set as 0. Even if we treat $\tau/0=\infty$, or view the model as a sequence of models with $\eta \to 0^+$, the three kernel-weighted moment functions still break down. When η becomes arbitrarily small, the fundamental strategy will return infinitesimal profit. It violates the assumption that the fundamental strategy beats the chartist strategy under arbitrarily deviation from $\Delta_{t-1}=0$, and invalidates $g_{2t}(\theta)$ and $g_{3t}(\theta)$, which were justified by arguing that the market is dominated by fundamentalists when G_1 occurs. Moreover, as a chartist ignores the fundamental value, α is also unidentified; thus $g_{4t}(\theta)$ is not well defined.

Given the difficulty of adapting it to the chartist-only scenario, we slightly modify the benchmark model. In a market with only chartists, the demand equation becomes $R_t(\theta) = \tau\Delta_{t-1}$. Notice that even without fundamentalists, the event G_1 remains well defined. Thus we introduce another kernel-weighted moment function

$$g_{9t}(\theta) = w_t^{G_1}(h_T)\left(|R_t^{\mathrm{r}}| - \tau\,|\Delta_{t-1}|\right),$$

where G_2 is replaced by G_1. This conditional moment is implied by the chartist-only model: when Δ_{t-1} is close to 0, the return must also be small.

When estimating the chartist-only model, we utilize $g_{9t}(\theta)$ along with the five unconditional moment functions $\{g_{jt}(\theta)\}_{j=1,5,6,7,8}$. The estimation involves six moments and two parameters (σ_μ,τ), so that the J-statistic still follows $\chi^2(4)$ asymptotically under the null hypothesis.Extension of the Model

S4 Extension of the Model

Homogeneity and time invariance of the conditional variance in returns is restrictive, especially during a financial crisis. In this section, we discuss the possibility of relaxing this assumption.

We define $\eta_{it} = 1/\left(A^f \cdot \text{var}_{it-1}^f[R_t^f]\right)$ to allow $\text{var}_{it-1}^f[R_t^f]$ to vary across i and t. Similarly, define $\tau_t = 1/\left(A^c \cdot \text{var}_{t-1}^c\left[R_t^c\right]\right)$, which is time-varying but individual invariant as the chartist strategy does not consider any private signal. It follows that for the fundamental strategy $\pi_{it}^f = \eta_{it}\left(\frac{\alpha\sigma_x}{1+\alpha}\left(\varepsilon_{it} - \delta_t\right)\right)^2$, and for the chartist strategy $\pi_t^c = \tau_t\Delta_{t-1}^2$. The investor chooses the fundamental strategy if $\pi_{it}^f \geq \pi_t^c$, and the demand of the risky asset is

$$q_{it}^* = q_{it}^{f*} \cdot \mathbf{1}\left\{\pi_{it}^f \geq \pi_t^c\right\} + q_{it}^{f*} \cdot \mathbf{1}\left\{\pi_{it}^f < \pi_t^c\right\}.$$

To compute the market aggregate demand, we need to specify the conditional variances since neither $\text{var}_{it-1}^f[R_t^f]$ nor $\text{var}_{t-1}^c\left[R_t^c\right]$ is observable from the data. A simple rule from the observed past history is an option for the chartist, while there is no consensus in the literature about the conditional variance of the fundamental strategy.

Consider imposing a parametric assumption on the joint distribution of $(\eta_{it}, \varepsilon_{it})$, for example, jointly normal i.i.d. across time. This simple specification introduces two extra parameters: the variance of η_{it} that captures the dispersion of the beliefs on the volatility, and the correlation coefficient between η_{it} and ε_{it}. Although the theoretical model can be simulated by the method of simulated moments (MSM), all the closed-forms in the aggregate demand and the thin-set identification are lost. Such difficulty arises even before we study any dynamic specification in (η_{it}), which will incur additional parameters.

Analysis becomes more tractable if we assume that the distribution of η_{it} and ε_{it} are independent. Let $\bar{\varepsilon}_{it}$ be the threshold such that $\pi_{it}^f = \pi_t^c$. We solve $\eta_{it}\left(\frac{\alpha\sigma_x}{1+\alpha}\left(\bar{\varepsilon}_{it} - \delta_t\right)\right)^2 = \tau_t\Delta_{t-1}^2$ to obtain

$$\bar{\varepsilon}_{it} = \delta_t \pm \frac{1+\alpha}{\alpha\sigma_x}\sqrt{\frac{\tau_t}{\eta_{it}}}\left|\Delta_{t-1}\right| = \delta_t \pm \zeta_{it-1},$$

where $\zeta_{it-1} = \frac{1+\alpha}{\alpha\sigma_x}\sqrt{\frac{\tau_t}{\eta_{it}}}\left|\Delta_{t-1}\right|$. Define the lower bound $\bar{\varepsilon}_{it}^m = \delta_t - \zeta_{it-1}$ and upper bound $\bar{\varepsilon}_{it}^M = \delta_t + \zeta_{it-1}$. As a result, the individual investment flow is

$$q_{it}^* = q_{it}^{f*} \cdot \mathbf{1}\left\{\varepsilon_{it} \in (-\infty, \bar{\varepsilon}_{it}^m] \cup [\bar{\varepsilon}_{it}^M, \infty)\right\} + q_t^{c*} \cdot \mathbf{1}\left\{\varepsilon_{it} \in (\bar{\varepsilon}_{it}^m, \bar{\varepsilon}_{it}^M)\right\}.$$

The probability of individual i adopting the chartist strategy is $m_{it} = \Lambda\left(\bar{\varepsilon}_{it}^{M}\right) - \Lambda\left(\bar{\varepsilon}_{it}^{m}\right)$, and the aggregate demand in the market is

$$\begin{aligned} D_t(\theta) &= \int_0^1 \int_{(-\infty,\bar{\varepsilon}_{it}^m]\cup[\bar{\varepsilon}_{it}^M,\infty)} \frac{\eta_{it}\alpha\sigma_x}{1+\alpha}\left(\varepsilon_{it}-\delta_t\right) d\Lambda\left(\varepsilon_{it}\right) di + \tau_t\Delta_{t-1}\int_0^1 m_{it} di \\ &= \frac{\alpha\sigma_x}{1+\alpha}\left(\int_0^1 \eta_{it}\left[\int_{-\infty}^{\bar{\varepsilon}_{it}^m} z d\Lambda(z) + \int_{\bar{\varepsilon}_{it}^M}^{\infty} z d\Lambda(z) - \left(1-m_{it}\right)\delta_t\right] di\right) + \tau_t\Delta_{t-1}\int_0^1 m_{it} di \\ &= \frac{\alpha\sigma_x}{1+\alpha}\left(\int_0^1 \eta_{it}\left[\varphi\left(\bar{\varepsilon}_{it}^m\right) - \varphi\left(\bar{\varepsilon}_{it}^M\right) - \left(1-m_{it}\right)\delta_t\right] di\right) + \tau_t\Delta_{t-1}\int_0^1 m_{it} di. \end{aligned} \quad \text{(S2)}$$

If we further assume $\tau_t = \tau/\varepsilon_t^c$ with ε_t^c being the proxy for $\mathrm{var}_{t-1}^c\left[R_t^c\right]$, we can pointly identify τ under the event G_2 as in the main text. This identified τ will depend on the choice of ε_t^c. On the other hand, if we assume $\eta_{it} = \eta/\varepsilon_{it}^v$, where ε_{it}^v is the shock to each individual's conditional variance independent of all other random variables, then the identification of (η,α) remains under the event G_1. Therefore, in this generalized model in which we allow time-varying and heterogeneous η_{it}, we are able to pointly identify the same parameter (η,α) as in the main text where a constant $\mathrm{var}_{t-1}^f\left[R_t^f\right]$ is assumed. As a result, the empirical estimates of (η,α) in the three periods in the main text are informative about the magnitude of these parameters.

If we start with the general model, nevertheless, MSM will be necessary to handle the integrals $\int_0^1 \eta_{it}\left[\varphi\left(\bar{\varepsilon}_{it}^m\right) - \varphi\left(\bar{\varepsilon}_{it}^M\right) - \left(1-m_{it}\right)\delta_t\right] di$ and $\int_0^1 m_{it} di$ in the demand equation (S2). We do not have simple closed-forms for these integrals as $\bar{\varepsilon}_{it}^m$, $\bar{\varepsilon}_{it}^M$ and m_{it} all depend on η_{it} and τ_t. Exploration of the conditional variance in this heterogeneous agent model would contribute to the theoretical modeling, asymptotic property of XMM-MSM, as well as the empirical findings. All these three aspects are new to the existing literature and they deserve thorough investigation in future research.

S5 Examples of Thin-Set Identification

Thin-set identification is not peculiar to our model. It is also useful for other heterogeneous agent models. Here we give two examples.

Example 1. Lux (1995) formalizes herd behavior in speculative markets in which bubbles emerge as self-organizing process of infection among traders. Let x be an index ranging from -1 (extremely pessimistic) to 1 (extremely optimistic). It characterizes the average opinion of speculative investors. The dynamics of x is governed by the differential equation

$$dx/dt = 2v\left(\tanh\left(ax\right) - x\cosh\left(ax\right)\right),$$

where a is a measure of the strength of herd behavior, and v is a variable for the speed of change. The fraction of optimistic trader is $0.5\left(x+1\right) \in [0,1]$ (Lux, 1995, pp.884–885). When $x = 1$, the fraction of optimistic trader is 1. □

Example 2. He and Westerhoff (2005) analyze the creation of bull or bear market via nonlinear interactions between market participants—consumers, producers and heterogeneous speculators—in a behavioral commodity market model. They model the market share of chartists as $1/\left(1 + d\left(F - S_t\right)^2\right)$, where d is a switching parameter, F is the long-run equilibrium price, and S_t is the commodity price at time t (He and Westerhoff, 2005, p.1582). When $F = S_t$, the fraction of chartists is 1. □

References

Antoine, B. and E. Renault (2012). Efficient minimum distance estimation with multiple rates of convergence. *Journal of Econometrics 170*(2), 350–367.

Chen, J., A. M. Variyath, and B. Abraham (2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics 17*(2), 426–443.

Fama, E. F. and K. R. French (2002). The equity premium. *The Journal of Finance 57*(2), 637–659.

Fan, J. and Q. Yao (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer.

Gagliardini, P., C. Gourieroux, and E. Renault (2011). Efficient derivative pricing by the extended method of moments. *Econometrica 79*(4), 1181–1232.

He, X.-Z. and F. H. Westerhoff (2005). Commodity markets, price limiters and speculative price dynamics. *Journal of Economic Dynamics and Control 29*(9), 1577–1596.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *The Annals of Statistics 25*(5), 2084–2102.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics 133*(1), 97–126.

Lux, T. (1995). Herd behaviour, bubbles and crashes. *The Economic Journal*, 881–896.

Qin, J. and J. Lawless (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics 22*, 300–325.