# Essays on Expectations and the Econometrics of Asset Pricing

Lof, Matthijs

May 2013

MATTHIJS LOF

# ESSAYS ON EXPECTATIONS AND THE ECONOMETRICS OF ASSET PRICING

# Acknowledgements

This thesis is the final result of more than a decade of studying econometrics, which began in the summer of 2000 at the University of Amsterdam. After completing my MSc, I took the rational next step of a traineeship at the European Central Bank. One year later, I found myself moving from Frankfurt to Helsinki, showing that financial markets are not alone in experiencing unpredictable events. After a couple of months in Finland, I considered returning to university and came across an open vacancy for a doctoral studentship in time-series econometrics at the University of Helsinki. With only days left to the application deadline, I scrambled together a research proposal on financial bubbles and crashes, inspired by the chaotic state of global markets at the time (October 2008). The final research presented in this thesis has in fact stayed remarkably close to this hastily prepared proposal. Following my application, I had an interview with Professors Markku Lanne and Pentti Saikkonen. A few days later, they offered me four years of funding (a rare luxury in academia), as well as the possibility to start the PhD program 'mid-season' in January 2009. Needless to say, I remain thankful to them for this opportunity.

Throughout all stages of writing this thesis, Markku Lanne has provided detailed and constructive suggestions, often within hours of receiving my work. Besides offering outstanding academic advice, Markku has provided invaluable help with respect to navigating the university's bureaucracy and dealing with the complexities of journal submissions and grant applications. The influence of Pentti Saikkonen has also been substantial. A short talk with Pentti is typically far more productive than weeks of reading and thinking on my own. When working with noncausal autoregressions in particular, these discussions helped me to truly understand the models and to frame my results.

# Contents

# Chapter 1

## Introduction

### 1.1   Expectations matter

Financial assets such as stocks or bonds can not be consumed or allocated for productive purposes. The only objective they serve is the reallocation of liquid funds over time. In exchange for an initial investment, the buyer of an asset receives a claim on future income in the form of cash flows paid by the corporation issuing the asset. The price of such an asset, whether it is a stock or a bond, should therefore be entirely determined by the expected present value of these cash flows, whether they are dividends or interest payments.

The idea that asset prices reflect expected future cash flows is both intuitive an appealing. Nevertheless, it constitutes one of the main puzzles in the field of asset pricing: Excess volatility. Stock prices are far more volatile than dividends. After the rational expectations revolution by Muth (1961) and Lucas (1972) swept through macroeconomics and finance, financial economists often assume in their models that investors take all available information into consideration in order to form optimal predictions regarding future dividends. A large body of empirical research (surveyed by Gilles and LeRoy, 1991) finds, however, that rational dividend expectations can not be sufficiently volatile to be the sole driver of price fluctuations.

In addition to dividend expectations, time-varying discount factors can contribute to price volatility. A claim on an expected payment of €100 in one year from now is in general worth

less than €100 today, for two reasons. First, the investor has to be compensated for not being able to access his invested money for one year. Second, the investor bears the risk that the issuer of the asset will be unable to pay the full amount of €100, or any amount at all, at the end of the year. The difference between the expected payoff and the price is in asset pricing models parameterized by a discount factor. If this discount factor varies over time, for example because the risk appetite of investors varies over time, prices could move without necessarily any news regarding future dividends. In recent decades, modeling the behavior of discount factors (alternatively: discount rates, state-price deflators, risk premiums) has been one of the main objectives of the asset pricing literature. As John Cochrane (2011) states, in his address to the American Finance Association:

> "Discount-rate variation is the central organizing question of current asset-pricing research. [...] Asset prices should equal expected discounted cash flows. Forty years ago, Eugene Fama (1970) argued that the expected part, "testing market efficiency," provided the framework for organizing asset-pricing research in that era. I argue that the "discounted" part better organizes our research today."

Although it is not unreasonable to assume that risk aversion, preferences and therefore discount factors change over time, I find Cochrane's claim that time-variation of the discount factor is the main or even the only relevant source of price fluctuations rather strong. To my judgment, there is certainly still scope for research on expectations. For one thing, it is an oversimplification to assume that all investors value assets according to expected dividends. Instead of buying an asset for its dividends, many investors make investments in the hope of short-term trading profits, thereby relying mainly on expectations on prices rather than dividends. Moreover, casual observation confirms that different investors may form different expectations. There would be little trade in a world of rational expectations and common knowledge (Lucas, 1978; Barberis and Thaler, 2003). The idea that speculative considerations can drive price fluctuations is not

new. For example, John Kenneth Galbraith (1961) notes, in his description of the run-up to the 1929 stock market crash:

> "At some point in the growth of a boom all aspects of property ownership become irrelevant except the prospect for an early rise in price. Income from the property, or even its long-run worth are now academic. [...] What is important is that tomorrow or next week market values will rise - as they did yesterday or last week - and a profit can be realized."

Nevertheless, many of the asset pricing models discussed by Cochrane (2011), are built around the concept of a rational representative agent, which leaves little to no room for speculative behavior or heterogeneous opinions to have an impact on prices.

Expectations matter. The essays in this thesis show that the way in which agents form expectations affects the dynamic properties of asset prices and therefore the appropriateness of different econometric tools used for empirical asset pricing. In addition to standard rational expectations models, I study the class of models introduced by Brock and Hommes (1997, 1998), in which boundedly rational agents may switch between various simple expectation rules. A crucial feature of these models is that not all agents have to follow the same expectation rule, but are allowed to form heterogeneous beliefs.

Chapters 2 and 3 present empirical estimations of two specific heterogeneous agent models. Since the data generating processes are assumed to be nonlinear, due to the agents' switching between expectation rules, I apply nonlinear regression models. The final two chapters deal with noncausal autoregressions. In Chapter 4, I show that noncausal autoregressions are better able than their causal counterparts to capture the dynamics of asset prices that are generated by heterogeneous agent models. Finally, in Chapter 5, I consider the estimation of a class of standard rational expectations models, and show that noncausality of the instruments does not necessarily have an impact on the consistency of the generalized method of moments (GMM) estimator.

This introductory chapter proceeds as follows. In Section 1.2, I describe the dataset of US aggregate stock prices, dividends and earnings, which is used throughout the essays in this thesis. Section 1.3 gives an overview of several univariate and multivariate time-series models, used for empirical asset pricing, with special focus on nonlinear and noncausal extensions of the benchmark autoregressive model. In Section 1.4, I review a small selection of asset pricing models. Section 1.5 provides summaries of the essays.

## 1.2    Stock prices, dividends and earnings

All empirical results presented in this thesis rely mainly on the same dataset of historical US stock prices, which is compiled, updated and published by Robert Shiller. The dataset contains monthly observations of the Standard & Poor's (S&P) 500 index, one of the prime stock market indices, constructed as a weighted average of the stock prices of 500 large publicly traded US companies. Although the S&P500 index was released only in 1957, Shiller has combined several data sources to construct a US stock market index going back all the way to 1871. Moreover, the dataset includes average dividends and earnings per share for the index. Detailed information on the sources and compilation of the index is found in Shiller (1989).

Figure 1.1 shows the level (price, $P_t$) of the index and the average dividends ($D_t$) and earnings ($E_t$) for the period 1871-2012. Due to exponential growth, these plots do not reveal much about price movements during the first 100 years. Rescaling the price by the level of the dividends, resulting in the price-dividend (PD) ratio, improves the picture a bit, although the peak experienced in the last 20-30 years still overshadows all previous fluctuations. This dominance is less profound for the price-earnings (PE) ratio. The peak around the millennium is clearly larger than in any period observed before, but the plot of the PE ratio also shows other interesting periods, such as the boom and bust around 1929 and the decreasing valuation during the 1970s. The difference in patterns of the PE and PD ratio is due to the fact that dividends as a fraction of earnings have been steadily declining over the last 60 years or so, which is depicted
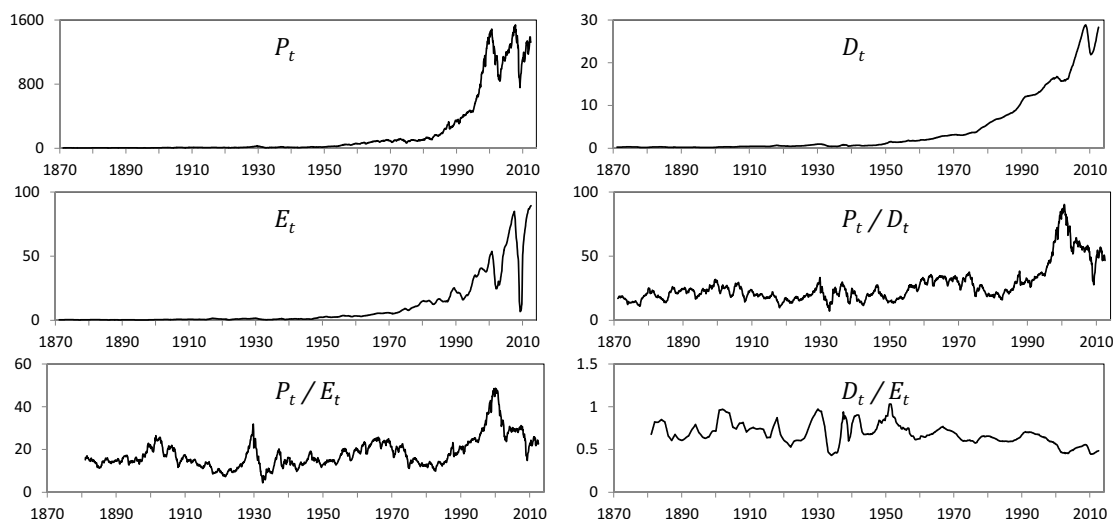
Figure 1.1: S&P 500 index ($P_t$), underlying dividends ($D_t$), earnings ($D_t$) and price-dividend (PD), price-earnings (PE) and dividend-earnings (DE) ratios. Monthly observations 01.1871-06.2012. For constructing the PE and DE ratio, earnings are smoothed over a period of 10 years, following the convention by Shiller (1989). Source: http://www.econ.yale.edu/~shiller/

in the final plot of Figure 1.1. Companies are distributing a declining share of their profits as dividends, which has resulted in higher PD ratios (Fama and French, 2001).

Financial economists are often interested in testing whether the (log) price is a random walk or, equivalently, whether log-differences (returns) are unpredictable white noise. Figure 1.2 shows annual, monthly and daily returns (left panel). The plotted time series show that returns are highly erratic and seem hard to predict. The autocorrelation plots in the middle panel, however, suggest that there is some degree of predictability, with significant first-order autocorrelations at the daily frequency and in particular at the monthly frequency. More evidence in favor of return predictability has been documented. In particular the PE ratio turns out to be a good predictor for returns (e.g. Campbell and Shiller, 2001, and Cochrane, 2011). Periods during which the S&P500 index is highly valued in terms of the PE ratio, are typically followed by low returns, while low valuations predict high returns over the next 5-10 years. This is evidence for mean reversion in stock prices, which contradicts the random walk assumption. High returns push up valuations, which in turn predicts low returns or decreasing valuations.

In addition to predictability of the level of returns, Figure 1.2 clearly shows dependence in the second moments of returns. The time series on the left show that extreme observations (regardless of the sign) typically occur within prolonged periods of high volatility, a phenomenon

5
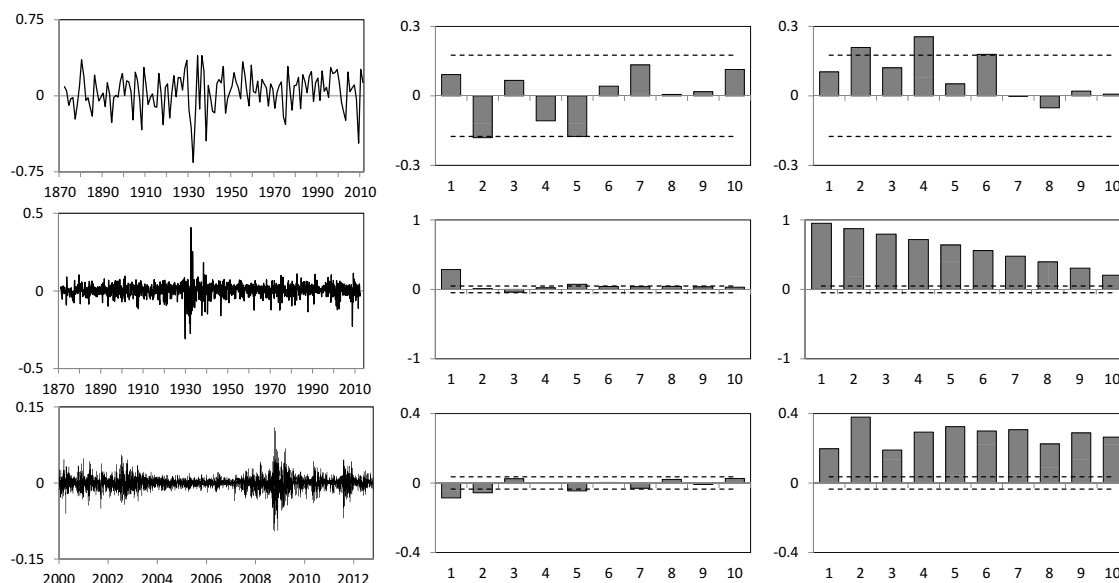
Figure 1.2: S&P 500 log-differences / returns. Annual observations 1871-2011 (top), monthly observations 01.1871-06.2012 (middle) and daily observations 4.1.2000-19.10.2012 (bottom). Autocorrelation plots for levels (middle) and squared returns (right), with 95% significance bounds. Sources: http://www.econ.yale.edu/~shiller/ and FRED® (Federal Reserve Economic Data)

referred to as volatility clustering. This becomes more evident from the plots on the right, which depict the autocorrelation functions of squared returns. In particular for higher frequencies, squared returns are highly autocorrelated.

Extreme returns are not only clustered, they occur rather often. Assuming a Gaussian distribution, absolute returns in deviation from the mean should exceed three standard deviations for only 0.1% of the observations. However, for the annual, monthly and daily data depicted in Figure 1.2, around 1.5% of the observations can in fact be classified as such extreme events. The distribution of returns therefore has 'fatter tails' than a Gaussian distribution. The fact that financial returns are non-Gaussian is well known (See e.g. Mandelbrot, 1963). Nevertheless, many theoretical asset pricing models are built on the assumption of Gaussianity (See e.g. Munk, 2013)

The observation that returns are clearly not white noise does not necessarily imply a rejection of the efficient market hypothesis, which states that prices should reflect all available information, thereby eliminating the possibility to achieve higher than average returns by making investment decisions based on publicly available information (Fama, 1970). Although there is evidence in favor of predictability over time for the aggregate stock market, it is a lot harder

6

to predict which specific stocks will outperform others. Although market inefficiencies have been documented (e.g. Gromb and Vayanos, 2010), many authors, including Malkiel (1973) and Fama and French (2010), evaluate historic returns achieved by institutional investors, to conclude that it is in fact very hard to create a portfolio in real time that is able to 'beat the market' for a prolonged period.

This thesis deals with stock prices only. The prices of many other financial assets, however, possess rather similar time-series properties. Figure 1.3 depicts daily observations of the US Dollar/Euro exchange rate, the yield on 10-year treasury bonds and the oil price over the period 4.1.2000-19.10.2012. Like with the S&P 500 index, these series show persistent, random-walk type behavior in levels, and strong volatility clustering in the returns.
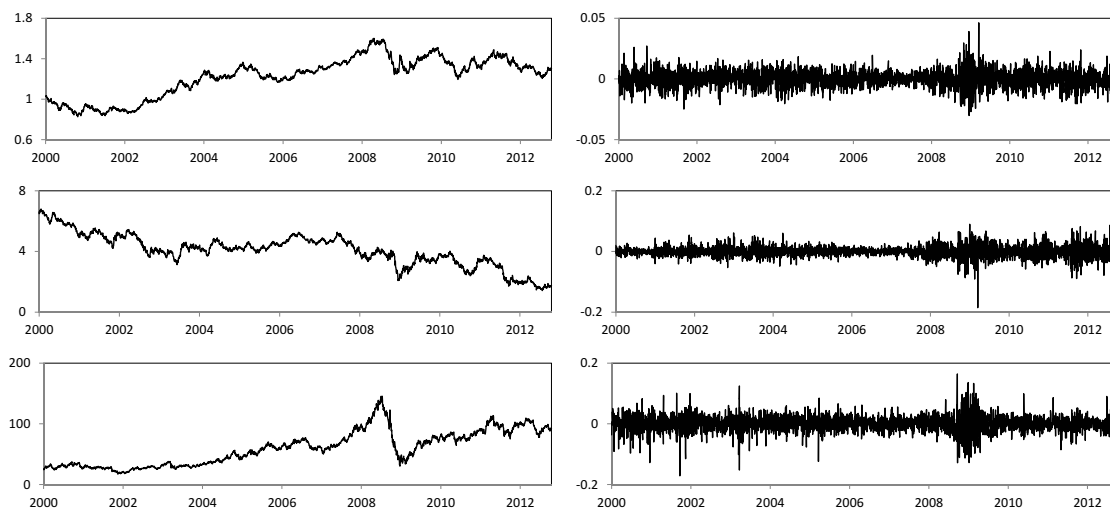


Figure 1.3: USD/EUR exchange rate (top), 10-year treasury yield (middle) and WTI crude oil price (bottom). Daily observations in levels (left) and log-differences / returns (right), 4.1.2000-19.10.2012. Sources: FRED® (Federal Reserve Economic Data)

## 1.3 Autoregressions

This section provides a brief overview of selected tools available to econometricians for analyzing time-series data. After outlining the benchmark autoregressive moving-average (ARMA) model, I discuss nonlinear and noncausal extensions.

A stationary time-series $y_t$, may be generated by the following ARMA($p,q$) process:

$$\alpha(L)y_t = \theta(L)\varepsilon_t, \qquad (1.1)$$

in which $\alpha(z) = 1 - \alpha_1 z - ... - \alpha_p z^p$, $\theta(z) = 1 + \theta_1 z + ... + \theta_q z^p$ and $\varepsilon_t \sim i.i.d.(0, \sigma^2)$ is an i.i.d. innovation, or error term. $L$ is a standard lag operator ($L^k x_t = x_{t-k}$). For example, an ARMA($1,1$) process takes the form:

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \qquad (1.2)$$

If $q = 0$, Equation (1.1) is referred to as an autoregressive AR($p$) process, while the restriction that $p = 0$ defines a moving-average MA($q$) process. Although Equation (1.1) may be supplemented with an intercept term, in this thesis I consider only zero-mean time series, in which case an intercept term becomes redundant.

If both polynomials $\alpha(z)$ and $\theta(z)$ have their roots outside the unit circle, the ARMA($p,q$) model has both infinite-order MA($\infty$) and AR($\infty$) representations:

$$
\begin{aligned}
MA(\infty): \qquad y_t &= \alpha(L)^{-1}\theta(L)\varepsilon_t \\
y_t &= \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}
\end{aligned}
$$

$$\qquad (1.3)$$

$$
\begin{aligned}
AR(\infty): \qquad \theta(L)^{-1}\alpha(L)y_t &= \varepsilon_t \\
y_t &= \sum_{j=1}^{\infty} \omega_j y_{t-j} + \varepsilon_t,
\end{aligned}
$$

in which $\sum_{j=0}^{\infty} \psi_j z^j = \psi(z) \equiv \alpha(z)^{-1}\theta(z)$ and $\sum_{j=0}^{\infty} \omega_j z^j = \omega(z) \equiv -\theta(z)^{-1}\alpha(z)$ (See Brockwell and Davis, 1991, for details). Since the ARMA($p,q$) process has an AR($\infty$) representation, it can sometimes be approximated quite well by a finite-order autoregressive AR($k$) process, with $k \geqslant p$:

$$
\begin{aligned}
\alpha(L)y_t &= \varepsilon_t \\
y_t &= \sum_{j=1}^{k} \alpha_j y_{t-j} + \varepsilon_t,
\end{aligned}
$$

$$\qquad (1.4)$$

Since the seminal contribution by Sims (1980), it has become a common approach in economics, at least for multivariate time series, to ignore moving-average terms and to consider pure autoregressions like (1.4), which in the case of multivariate time-series is referred to as a vector autoregression (VAR). In this thesis, I follow this convention also for univariate time-series. One reason for omitting the moving average terms is the simplicity of estimation. For any observed time series $y_t$, an AR($k$) model can be estimated consistently by regressing $y_t$ on its own $k$ lags, using ordinary least squares (OLS). Another reason is that the AR($k$) process (1.4) is nested in the nonlinear and noncausal autoregressions discussed below. Avoiding moving-average terms therefore facilitates a more clear comparison between the different models applied in this thesis.

Nonlinearity is not a well-defined concept (See, for example, the discussion in Teräsvirta et al., 2010, Chapter 1). One way to think about nonlinearity in the context of autoregressive models, following Granger (2008), is to allow for time-varying parameters:

$$\alpha_t(L)y_t \;\; = \;\; \varepsilon_t, \tag{1.5}$$

in which $\alpha_t(z) = 1 - \alpha_{1,t}z - ... - \alpha_{k,t}z^k$. The parameters $\alpha_{i,t}$ (for $i = 1,...,k$) vary over time following some stochastic or deterministic process. A well known parametric example is the smooth transition autoregressive (STAR) model:

$$(\gamma(L)(1-G_t) + \delta(L)G_t)y_t \;\; = \;\; \varepsilon_t, \tag{1.6}$$

in which the two autoregressive polynomials $\gamma(z)$ and $\delta(z)$ define the regimes of the model, while the transition function $G_t$ determines the weights of each regime. The STAR model (1.6) corresponds to the time-varying parameter model (1.5) such that the time-varying parameters are in fact a time-varying weighted average of two constant parameters: $\alpha_{i,t} = \gamma_i(1 - G_t) + \delta_i G_t$ (for $i = 1,...,k$). In the case that $\gamma(z) = \delta(z)$, the STAR model reduces to the linear autoregression (1.4). In the STAR models considered in this thesis, the transition function

$G_t$ takes the form of a logistic function:

$$G_t \;=\; (1 + \exp[-\beta(s_t - c)])^{-1}, \tag{1.7}$$

in which case (1.6) defines a Logistic STAR (LSTAR) model. In this case, the transition between regimes depends on a constant parameter $c$, a slope parameter $\beta$ and a transition variable $s_t$. The slope parameter $\beta$ determines the smoothness of the transitions. If $0 < \beta < \infty$ the transition function fluctuates smoothly over the interval $0 < G_t < 1$. If $\beta = 0$, the transition function is constant and the STAR (1.6) reduces to the linear autoregression (1.4), with $\alpha_i = \dfrac{\gamma_i + \delta_i}{2}$ (for $i = 1,...,k$). If $\gamma = \infty$, $G_t$ is in each period either zero or one, depending whether $s_t$ is smaller or larger than $c$. In this case, the STAR is actually a Threshold Autoregressive (TAR) model. The transition variable is typically a lagged value of the endogenous variable: $s_t = y_{t-d}$ (for $d > 0$), but it can be any exogenous or predetermined variable. In Chapter 2 of this thesis, I consider the case where $s_t$ is a linear combination of multiple predetermined variables. As long as the transition variables are predetermined or exogenous, the STAR model can be estimated consistently by nonlinear least squares (NLS) or maximum likelihood (ML).

Various alternatives to the benchmark STAR model have been considered in the literature. For example, instead of a logistic function, the transition function may also be an exponential function, resulting in the Exponential STAR (ESTAR) model. Several other extensions, including multivariate and multiple-regime alternatives, as well as details on estimating STAR models are discussed by Teräsvirta (1994), Van Dijk et al. (2002) and Teräsvirta et al. (2010).

Returning to linear autoregressions, I assumed above that the polynomial $\alpha(z)$ in (1.4) has its roots outside the unit circle. If instead, one or more of the roots lie on the unit circle (unit root), the AR process (1.4) is nonstationary. In this thesis, I consider stationary time-series only, which is in some cases established by differencing the variables. A third case, which has so far hardly been considered in economic applications, is that one or more of the roots of $\alpha(z)$ lie inside the unit circle. In this case, (1.4) defines a noncausal autoregression (Brockwell and

Davis, 1991). Lanne and Saikkonen (2011b) recently introduced a novel parameterization of the noncausal AR($k$) process (depending on $k$ lags), to a 'forward-looking' noncausal AR($r,s$) process depending on $r$ lags as well as $s$ leads, with $r+s=k$:

$$\phi(L)\varphi(L^{-1})y_t = \varepsilon_t, \tag{1.8}$$

with $\phi(L) = 1 - \phi_1 L - ...\phi_r L^r$, $\varphi(L^{-1}) = 1 - \varphi_1 L^{-1} - ...\varphi_r L^{-s}$. Both polynomials have their roots outside the unit circle. If $\varphi_j \neq 0$, for some $j \in \{1,..,s\}$, (1.8) is a noncausal process, which may be referred to as purely noncausal if $\phi_1 = ... = \phi_p = 0$. When $y_t$ is a vector, (1.8) defines a noncausal vector autoregressive process VAR($r,s$) (Lanne and Saikkonen, 2013). An interesting feature of the noncausal AR($r,s$) process is its MA representation, which is both backward- and forward-looking:

$$y_t \;\; = \;\; \varphi(L^{-1})^{-1}\phi(L)^{-1}\varepsilon_t \;\; = \;\; \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \tag{1.9}$$

Since current observations of $y_t$ depend on future values of $\varepsilon_t$, it is no longer appropriate to refer to $\varepsilon_t$ as an innovation. One way of interpreting noncausality is that the time series $y_t$ is generated by an economy in which agents form expectations based on information that is unobservable to an econometrician who observes realizations of $y_t$ only. The residuals $\varepsilon_t$ can therefore not be interpreted as true shocks to the agents information set, i.e. they are nonfundamental (Hansen and Sargent, 1991). In Chapter 4, I simulate examples of nonfundamentalness, by generating time series which are part of a multivariate (VAR) or nonlinear (STAR) model. I then consider an econometrician who observes one of these time series without knowledge of the correct data generating process and tries to fit a linear univariate autoregression to the data. In many cases, due to the missing information, noncausal autoregressions provide the best fit.

In order to estimate a noncausal autoregression, $\varepsilon_t$ has to be non-Gaussian. For any noncausal autoregression, a causal autoregression with identical first- and second-order moments can be found, which can not be distinguished from its noncausal counterpart if $\varepsilon_t$ is Gaussian. Lanne and Saikkonen (2011b, 2013) provide ML estimators for noncausal (V)ARs under the assumption of t-distributed errors. This is typically not a troubling assumption in the case of

macroeconomic and financial data, as the t-distribution captures the fat tails discussed in the previous section better than the Gaussian distribution.

Besides the autoregressive polynomial, it is also possible to allow for the moving-average polynomial in (1.1) to have its roots inside the unit circle, resulting in a noninvertible ARMA process (Meitz and Saikkonen, 2013). Although I do not use this class of models in this thesis, it is worth mentioning in particular since these models have been proven useful recently in testing for predictability of stock returns (Lanne et al., 2013).

To capture the observed volatility clustering as described in the previous section, another class of nonlinear models is often used in financial econometrics. The assumption that the error term $\varepsilon_t$ is *i.i.d.*, with a constant conditional variance, is replaced by the assumption that the conditional variance of $\varepsilon_t$ varies over time, depending on lagged squared error terms, resulting in the so-called autoregressive conditional heteroscedasticity (ARCH($p$)) model, introduced by Engle (1982):

$$
\begin{aligned}
\varepsilon_t &= \sigma_t v_t \\
\sigma_t^2 &\equiv E_t\left[\varepsilon_t^2\right] = \rho_0 + \rho(L)\varepsilon_t^2,
\end{aligned}
\tag{1.10}
$$

in which $\rho(z) = 1 + \rho_1 z + ... + \rho_p z^p$. As volatility is often fairly persistent, a high value of $p$ can be required to obtain a satisfactory fit. Bollerslev (1986) therefore introduced the Generalized ARCH (GARCH($p,q$)) model:

$$
\begin{aligned}
\varepsilon_t &= \sigma_t v_t \\
\sigma_t^2 &\equiv E_t\left[\varepsilon_t^2\right] = \rho_0 + \rho(L)\varepsilon_t^2 + \delta(L)\sigma_t^2,
\end{aligned}
\tag{1.11}
$$

in which $\delta(z) = 1 + \delta_1 z + ... + \delta_q z^q$. A GARCH(1,1) is often able to capture rather persistent volatility and is therefore preferred to the less parsimonious ARCH($p$) model with a high number of lags $p$. (G)ARCH models are mainly useful in modeling volatility at high frequencies. Throughout this thesis, as I am dealing with low-frequency data only, I assume that $\varepsilon_t$ is *i.i.d.* and therefore do not consider (G)ARCH type specifications.

## 1.4 Discount factors, rationality and heterogeneity

I introduce the linear present value model and go over two generalizations that are considered in this thesis: Consumption-based asset pricing and boundedly rational heterogeneous expectations. The price ($P_t$) of an asset should equal the discounted sum of the expected price in the next period and any expected dividends ($D_{t+1}$) paid out in the meantime:

$$P_t = \delta E_t \left[ P_{t+1} + D_{t+1} \right]. \tag{1.12}$$

Iterating this equation forward results in the present value model, in which the price is determined by discounted dividend expectations only:

$$P_t = \sum_{i=1}^{\infty} \delta^i E_t \left[ D_{t+i} \right]. \tag{1.13}$$

LeRoy and Porter (1981) and Shiller (1981) test the present value model by analytically deriving bounds for the volatility prices, implied by the present value model (1.13) and observed dividends. The observation that these bounds are violated by the volatility of observed prices is referred to as excess volatility. The result of excess volatility is robust to several alternative tests (e.g. Campbell and Shiller, 1987, 1988 and West, 1988), typically involving a vector autoregressive representation of prices and dividends.

Partly motivated by the rejection of linear present value models, asset pricing research has moved largely towards time-varying discount factors (See the surveys by Campbell, 2000, and Cochrane, 2011):

$$P_t = E_t \left[ \zeta_{t+1} \left( P_{t+1} + D_{t+1} \right) \right], \tag{1.14}$$

in which $\zeta_t$ denotes the stochastic discount factor (SDF), which varies over time according to a certain stochastic process. A popular approach is the consumption-based discount factor, linking asset markets to the real economy. The idea is that in each period, a representative agent faces a choice between consuming the entire allocation of wealth, or postponing consumption

by investing part of the wealth in a financial asset. The optimal discount factor for valuing the asset can be shown to equal the intertemporal marginal rate of substitution:

$$\zeta_{t+1} \;\; = \;\; \delta \frac{U'(C_{t+1})}{U'(C_t)}, \tag{1.15}$$

in which $U'(\cdot)$ is the marginal utility of consumption, i.e. the first derivative of the utility function $U(\cdot)$ (See, e.g. Rubinstein, 1976, Lucas, 1978, Campbell, 2003, for details). Hansen and Singleton (1982) show that this model can be estimated by the generalized method of moments (GMM), using data on returns and aggregate consumption. The assumption of rational expectations means that the difference between the expectation and realization is orthogonal to all observable information. Equation (1.14) therefore implies the following moment condition:

$$E\left[\left(\zeta_{t+1}R_{t+1} - 1\right)\mathbf{z}_{t-1}\right] = 0. \tag{1.16}$$

in which $R_{t+1} = \dfrac{P_{t+1} + D_{t+1}}{P_t}$ and $\mathbf{z}_{t-1}$ is a vector of predetermined instruments. Hansen and Singleton (1982) choose lagged values of returns and consumption as instruments and assume a constant relative-risk aversion utility function $(U(C_t) = (1 - \gamma)^{-1} C_t^{1-\gamma})$. When the risk aversion coefficient $\gamma$ is equal to zero, the utility function is linear implying that agents are risk-neutral and the SDF (1.15) becomes constant as in (1.12). I return to this procedure in Chapter 5.

Although the SDF can account for additional volatility in asset prices, the consumption-based approach creates its own empirical problems such as the equity premium puzzle (Mehra and Prescott, 1985). Observed stock returns are rather high, which implies an unrealistically high degree of risk aversion $\gamma$. To overcome this problem, various more complex utility functions have been proposed in order to generate high returns for moderate values of $\gamma$ (e.g. Epstein and Zin, 1989, and Campbell and Cochrane, 1999).

Besides time-variation in the discount factor, it is also possible to allow for time-variation in the expectation operator. In the present value model (1.12) and the SDF model (1.14), expectations are assumed to be rational. Instead, several behavioral finance models have been proposed

in which expectations are non-rational, based on limited information sets, and possibly hetero-geneous (e.g. De Long et al., 1990a,b, Barberis et al., 1998, or Hong and Stein, 1999). In Chapters 2, 3 and 4, I consider the class of models proposed by Brock and Hommes (1997, 1998), in which assets are priced by $H$ types of boundedly rational agents who are allowed to form heterogeneous expectations:

$$P_t = \delta \sum_{h=1}^{H} G_{h,t} E_t^h [P_{t+1} + D_{t+1}], \tag{1.17}$$

in which $G_{h,t}$ is the fraction of agents forming expectations according to $E_t^h [\cdot]$ at time $t$. Brock and Hommes (1997, 1998) assume that the expectation operator $E_t^h [\cdot]$ is a simple linear uni-variate prediction rule, not necessarily taking into account all available information. Agents are allowed to switch between prediction rules, or strategies, based on evolutionary considera-tions: More successful strategies become more popular. To this end, the fraction of each type is modeled by multinomial logit probabilities:

$$G_{h,t} = \frac{\exp \left[ \beta U_{h,t-1} \right]}{\displaystyle\sum_{i=1}^{H} \exp \left[ \beta U_{i,t-1} \right]}, \tag{1.18}$$

in which $U_{h,t}$ is some metric evaluating the past performance of strategy $h$, such as realized trading profits or forecast accuracy. In Chapter 2, I consider a variant of this model in which the fractions of agents are determined by macroeconomic conditions. The metric $U_{h,t}$ is therefore replaced by a set of macroeconomic variables. Depending on the specification of the prediction rules, the heterogeneous agent model (1.17)-(1.18) may be represented by a STAR model like (1.6)-(1.7), such that the different regimes represent different prediction rules. Using annual data on the S&P500 index, Boswijk et al. (2007) estimate a specific two-type example ($H = 2$) of model (1.17)-(1.18), which is discussed in detail in Chapter 2.

## 1.5   Review of the essays

### *Chapter 2: Heterogeneity in stock prices:*
### *A STAR model with multivariate transition function*

A heterogeneous agent asset pricing model, featuring fundamentalists and chartists, is applied to the price-dividend and price-earnings ratios of the S&P500 index. Agents update their beliefs according to macroeconomic information, as an alternative to the evolutionary selection scheme in the heterogeneous agent models by Brock and Hommes (1998).

The asset pricing model can be parametrized as a STAR model, in which the two autoregressive regimes represent the beliefs of each type of agent. To facilitate regime-switching based on macroeconomic conditions, I generalize the transition function of the univariate STAR model to a multivariate transition function, and propose a procedure based on linearity testing, following Luukkonen et al. (1988), to select the appropriate linear combination of transition variables from a larger set of macroeconomic variables. The results indicate that during periods of favorable economic conditions the fraction of chartists increases, causing stock prices to decouple from fundamentals.

### *Chapter 3: Rational speculators, contrarians and excess volatility*

In Chapter 3, I consider an evolutionary asset pricing model with three types of agents. Besides rational long-term investors, that value assets according to expected long-term dividends, the model includes rational and contrarian speculators with shorter investment horizons. In contrast to Chapter 2, in which the agents choose between simple univariate expectation rules, in this chapter the expectations of all agents are anchored in the same VAR model, which implies that the VAR approach for testing present value models (Campbell and Shiller, 1987, 1988) can be applied to evaluate the model empirically.

Supplementing the standard present value model with speculative agents dramatically improves the model's ability to replicate the observed dynamics of US stock prices over the period 1871-2011. In particular the existence of contrarians can explain some of the most volatile

episodes including the 1990s bubble, suggesting this was not a rational bubble. After allowing for heterogeneous expectations, there is little evidence for time-variation in the discount factor.

## *Chapter 4: Noncausality and asset pricing*

Recent literature finds that many macroeconomic and financial variables are noncausal, in the sense that, within the class of linear (vector) autoregressions, these variables are best described by noncausal models. In Chapter 4, I show that US stock prices are also noncausal. This implies that agents' expectations are not revealed to an outside observer such as an econometrician observing only realized market data.

I show by simulation that misspecification of agents' information sets or expectation formation mechanisms may lead to noncausal autoregressive representations. In particular, asset prices are found to be noncausal when the data are generated by heterogeneous agent models of the type considered by Brock and Hommes (1998).

## *Chapter 5: GMM estimation with noncausal instruments under rational expectations*

I depart from the assumption of bounded rationality in Chapter 5, and consider a class of rational expectations models, of which the standard consumption-based asset pricing model is a specific example.

Lanne and Saikkonen (2011a) show that the GMM estimator is inconsistent, when the instruments are lags of variables that admit a noncausal autoregressive representation. I argue that this inconsistency depends on the distributional assumption that the error terms in the regression model and in the noncausal autoregressive representation are jointly *i.i.d.*, which does not always hold. In particular under the assumption of rational expectations, which is the identifying assumption for many macroeconomic and financial applications of GMM (e.g. Hansen and Singleton, 1982), the GMM estimator is found to be consistent. This result is derived in a linear context and illustrated by simulation of a nonlinear asset pricing model.

# References

Barberis, N., A. Shleifer, and R. Vishny: 1998, 'A model of investor sentiment'. *Journal of financial economics* **49**(3), 307–343.

Barberis, N. and R. Thaler: 2003, 'A survey of behavioral finance'. *Handbook of the Economics of Finance* **1**, 1053–1128.

Bollerslev, T.: 1986, 'Generalized autoregressive conditional heteroskedasticity'. *Journal of econometrics* **31**(3), 307–327.

Boswijk, H. P., C. H. Hommes, and S. Manzan: 2007, 'Behavioral heterogeneity in stock prices'. *Journal of Economic Dynamics and Control* **31**(6), 1938–1970.

Brock, W. A. and C. H. Hommes: 1997, 'A Rational Route to Randomness'. *Econometrica* **65**(5), 1059–1096.

Brock, W. A. and C. H. Hommes: 1998, 'Heterogeneous beliefs and routes to chaos in a simple asset pricing model'. *Journal of Economic Dynamics and Control* **22**(8-9), 1235–1274.

Brockwell, P. J. and R. A. Davis: 1991, *Time Series: Theory and Methods, Second Edition*. New York, NY: Springer-Verlag, 1991 edition.

Campbell, J.: 2000, 'Asset pricing at the millennium'. *The Journal of Finance* **55**(4), 1515–1567.

Campbell, J.: 2003, 'Consumption-based asset pricing'. In: G. Constantinides, M. Harris, and R. Stulz (eds.): *Handbook of the Economics of Finance*, Vol. 1B. Elsevier, Chapt. 13.

Campbell, J. and J. Cochrane: 1999, 'By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market'. *The Journal of Political Economy* **107**(2), 205–251.

Campbell, J. and R. Shiller: 1987, 'Cointegration and Tests of Present Value Models'. *Journal of Political Economy* **95**(5), 1062–1088.

Campbell, J. Y. and R. J. Shiller: 1988, 'The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors'. *Review of Financial Studies* **1**(3), 195–228.

Campbell, J. Y. and R. J. Shiller: 2001, 'Valuation Ratios and the Long-Run Stock Market Outlook: An Update'. *NBER Working Papers* (8221).

Cochrane, J. H.: 2011, 'Presidential Address: Discount Rates'. *The Journal of Finance* **66**(4), 1047–1108.

De Long, J., A. Shleifer, L. Summers, and R. Waldmann: 1990a, 'Noise Trader Risk in Financial Markets'. *Journal of Political Economy* **98**(4), 703–738.

De Long, J., A. Shleifer, L. Summers, and R. Waldmann: 1990b, 'Positive Feedback Investment Strategies and Destabilizing Rational Speculation'. *The Journal of Finance* pp. 379–395.

Engle, R.: 1982, 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation'. *Econometrica: Journal of the Econometric Society* pp. 987–1007.

Epstein, L. and S. Zin: 1989, 'Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework'. *Econometrica* **57**(4), 937–969.

Fama, E.: 1970, 'Efficient Capital Markets: A Review of Theory and Empirical Work'. *The Journal of Finance* **25**(2), 383–417.

Fama, E. and K. French: 2010, 'Luck versus Skill in the Cross-Section of Mutual Fund Returns'. *The Journal of Finance* **65**(5), 1915–1947.

Fama, E. F. and K. R. French: 2001, 'Disappearing dividends: changing firm characteristics or lower propensity to pay?'. *Journal of Financial Economics* **60**(1), 3–43.

Galbraith, J.: 1961, *The Great Crasch 1929*. Middlesex, England. Penguin Books.

Gilles, C. and S. LeRoy: 1991, 'Econometric aspects of the variance-bounds tests: A survey'. *Review of Financial Studies* **4**(4), 753–791.

Granger, C.: 2008, 'Non-linear models: Where do we go next-Time varying parameter models?'. *Studies in Nonlinear Dynamics & Econometrics* **12**(3).

Gromb, D. and D. Vayanos: 2010, 'Limits of arbitrage'. *Annual review of financial economics* **2**, 251–275.

Hansen, L. P. and T. J. Sargent: 1991, 'Two Difficulties in Interpreting Vector Autoregressions'. In: L. P. Hansen and T. J. Sargent (eds.): *Rational Expectations Econometrics*. Westview Press, Inc., Boulder, CO, pp. 77–119.

Hansen, L. P. and K. J. Singleton: 1982, 'Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models'. *Econometrica* **50**(5), 1269–86.

Hong, H. and J. Stein: 1999, 'A unified theory of underreaction, momentum trading, and overreaction in asset markets'. *The Journal of Finance* **54**(6), 2143–2184.

Lanne, M., M. Meitz, and P. Saikkonen: 2013, 'Testing for Predictability in a Noninvertible ARMA Model'. *Journal of Financial Econometrics* (forthcoming).

Lanne, M. and P. Saikkonen: 2011a, 'GMM Estimation with Noncausal Instruments'. *Oxford Bulletin of Economics and Statistics* **73**(5), 581–592.

Lanne, M. and P. Saikkonen: 2011b, 'Noncausal Autoregressions for Economic Time Series'. *Journal of Time Series Econometrics* **3**(3), Article 2.

Lanne, M. and P. Saikkonen: 2013, 'Noncausal vector autoregression'. *Econometric Theory* (forthcoming).

LeRoy, S. F. and R. D. Porter: 1981, 'The Present-Value Relation: Tests Based on Implied Variance Bounds'. *Econometrica* **49**(3), 555–74.

Lucas, R.: 1972, 'Expectations and the Neutrality of Money'. *Journal of economic theory* **4**(2), 103–124.

Lucas, R. E. J.: 1978, 'Asset Prices in an Exchange Economy'. *Econometrica* **46**(6), 1429–45.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta: 1988, 'Testing linearity against smooth transition autoregressive models'. *Biometrika* **75**(3), 491–499.

Malkiel, B.: 1973, *A Random Walk Down Wall Street*. W. W. Norton.

Mandelbrot, B.: 1963, 'The variation of certain speculative prices'. *Journal of business* pp. 394–419.

Mehra, R. and E. Prescott: 1985, 'The equity premium: A puzzle'. *Journal of monetary Economics* **15**(2), 145–161.

Meitz, M. and P. Saikkonen: 2013, 'Maximum likelihood estimation of a noninvertible ARMA model with autoregressive conditional heteroskedasticity'. *Journal of Multivariate Analysis* **114**, 227 – 255.

Munk, C.: 2013, *Financial Asset Pricing Theory*. Oxford University Press (forthcoming).

Muth, J.: 1961, 'Rational expectations and the theory of price movements'. *Econometrica: Journal of the Econometric Society* pp. 315–335.

Rubinstein, M.: 1976, 'The Valuation of Uncertain Income Streams and the Pricing of Options'. *The Bell Journal of Economics* pp. 407–425.

Shiller, R.: 1989, *Market volatility*. MIT Press.

Shiller, R. J.: 1981, 'Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?'. *American Economic Review* **71**(3), 421–36.

Sims, C.: 1980, 'Macroeconomics and Reality'. *Econometrica* **48**(1), 1–48.

Teräsvirta, T.: 1994, 'Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models'. *Journal of the American Statistical Association* **89**(425), 208–218.

Teräsvirta, T., D. Tjostheim, and C. W. J. Granger: 2010, *Modelling Nonlinear Economic Time Series*. Oxford University Press.

Van Dijk, D., T. Teräsvirta, and P. Franses: 2002, 'Smooth transition autoregressive models: A survey of recent developments'. *Econometric Reviews* **21**(1), 1–47.

West, K. D.: 1988, 'Dividend Innovations and Stock Price Volatility'. *Econometrica* **56**(1), 37–61.

# Chapter 2

## Heterogeneity in stock prices:
## A STAR model with multivariate transition function[1]

## 2.1  Introduction

Linear asset pricing models based on the efficient market hypothesis (EMH) are not well suited to explain the observed dynamics of financial markets. According to these models, asset prices reflect a rational forecast by the market of future cash flows (dividends) generated by the asset and are therefore expected to be smoother than the actual cash flows. However, financial asset prices such as stock prices are historically more volatile than real economic activity including corporate earnings and dividends. Several studies (e.g. LeRoy and Porter, 1981; Shiller, 1981; West, 1988; Campbell and Shiller, 1988, 2001) discuss this excess volatility in financial markets and conclude that stock prices can not be explained by expected dividends alone.

Heterogeneous agent models provide an alternative to the EMH. In these models, the single representative rational agent is replaced by boundedly rational agents who are heterogeneous in beliefs, are not necessarily forecasting future dividends and may switch between trading strategies over time. Hommes (2006) and Manzan (2009) provide surveys of such models in economics and finance. The model in this paper is based on the work by Brock and Hommes (1997, 1998), who introduce a simple analytically tractable heterogeneous agent model with two types of agents: Fundamentalists and chartists. Fundamentalists believe, in accordance with the EMH, that asset prices will adjust toward their fundamental value. Chartists (or trend-

---

[1]This chapter is based on an article published in the *Journal of Economic Dynamics and Control* (Lof, 2012)

followers) speculate on the persistence of deviations from the fundamental value. I use data on the S&P500 index to estimate a heterogeneous agent model in which macroeconomic and financial variables simultaneously govern the agents' switching between strategies. It turns out that during periods of high economic growth, agents switch from fundamentalism to chartism, i.e. loose sight of fundamentals and become more interested in following recent trends in asset prices, which causes asset price bubbles to inflate.

Heterogeneous agent models are typically estimated empirically using regime-switching regression models, with the distinct regimes representing the expected asset pricing processes according to each type of agent. In particular smooth-transition regime-switching models such as the smooth-transition autoregressive (STAR) models (Teräsvirta, 1994) are suitable, as the modeled process is a time-varying weighted average of the distinct regimes. The time-varying weights of the regimes are then interpretable as the fractions of agents belonging to each type.

Recent studies have estimated asset pricing models featuring chartists and fundamentalists for several types of asset prices including exchange rates (Manzan and Westerhoff, 2007; De Jong et al., 2010), option prices (Frijns et al., 2010), oil prices (Reitz and Slopek, 2009; Ter Ellen and Zwinkels, 2010) and other commodity prices (Reitz and Westerhoff, 2007). Boswijk et al. (2007) apply the model by Brock and Hommes (1998) to price-dividend (PD) and price-earnings (PE) ratios of the US stock market, finding that the unprecedented stock valuations observed during the 1990s are the result of a prolonged dominant position of the chartist type over the fundamentalist type.

Agents are in general assumed to switch between strategies based on evolutionary considerations. Boswijk et al. (2007) follow Brock and Hommes (1998) by letting the agents choose their regime based on the realized profits of each type. Alternatively, the switching may be based on relative forecast errors (Ter Ellen and Zwinkels, 2010), or on the distance between the actual and fundamental price (Manzan and Westerhoff, 2007). In this paper, the agents' choice of strategy is not evolutionary, but varies instead over the business cycle. In practice, this means I estimate a STAR model, in which the transition function depends on a linear combination of exogenous or predetermined macroeconomic variables. This framework allows for identifying

22

the macroeconomic conditions under which chartism or fundamentalism dominates the market.

The result that chartism is associated with economic expansion is novel but can be related to existing results in the literature on the effects of the real economy on financial markets. For example, Fama and French (1989), Campbell (2003) and Cooper and Priestley (2009), amongst others, study the variation of risk aversion over the business cycle, and find more risk appetite on financial markets during economic upturns. The interpretation of countercyclical risk premiums is different from this paper. Instead of a rational representative agent becoming less risk averse, I assume that under favorable economic conditions an increasing fraction of agents chooses a more speculative trading strategy by becoming chartist. These findings are, however, not necessarily inconsistent, as chartists are sometimes described as being less risk averse than fundamentalists (Chiarella and He, 2002; Chiarella et al., 2009). Using a cross-section of US stock returns, Chordia and Shivakumar (2002) find that momentum strategies are profitable only during the most expansionary periods of the business cycle. Without making any agent-based interpretations, Spierdijk et al. (2012) use a panel of stock market indices from 18 OECD countries to find that the speed of mean reversion towards the fundamental value accelerates during periods of high economic uncertainty. This result confirms my findings since a high speed of mean reversion implies a high fraction of fundamentalists.

The STAR model is typically univariate, in which the transition between regimes depends on a lag of the dependent variable as in Teräsvirta (1994). Alternatively, the transition function may depend on a single exogenous or predetermined transition variable as in Reitz and Westerhoff (2003), Reitz and Taylor (2008) and Reitz et al. (2011), who study the nonlinear effects of purchasing power parity and central bank policies on exchange rates. In contrast to these studies, I allow for a multivariate transition function depending on multiple exogenous or predetermined transition variables with unknown weights, in order to estimate the nonlinear effects of multiple economic variables simultaneously. Estimating this multivariate STAR model raises two difficulties compared to the univariate STAR: Selection of the transition variables to include, and estimation of their weights. Medeiros and Veiga (2005) and Becker and Osborn (2012) consider estimating STAR models with unknown weighted sums of transition

variables, but both are limited to univariate models in which the transition functions depend on linear combinations of different lags of the dependent variable. I propose to apply the linearity test by Luukkonen et al. (1988) to select the transition variables from a large set of information and simultaneously estimate their respective weights in the transition function. The resulting STAR model with multivariate transition function provides a better fit to the PD and PE ratios than linear models and STAR models with a single transition variable do, while the estimates support the idea of a smooth transition between chartism and fundamentalism.

The next section presents the heterogeneous agent model and the STAR specification in more detail. Data descriptions and linearity tests are given in Section 2.3 while Section 2.4 presents estimation results, interpretation and diagnostic checks. Section 2.5 concludes.

## 2.2   The model

In a simple linear present value asset pricing model, consistent with the efficient market hypothesis, the price of a financial asset ($P_t$) equals the discounted sum of the expected asset price next period and any expected cash flows (dividends, $D_{t+1}$) paid out on the asset in the coming period (Gordon, 1959). Iterating forward, the price can be expressed as a infinite sum of discounted expected dividends:

$$P_t = \frac{1}{1+r}E_t\left[P_{t+1}+D_{t+1}\right]=\sum_{i=1}^{\infty}\frac{1}{(1+r)^i}E_t\left[D_{t+i}\right], \tag{2.1}$$

in which the constant discount factor is given by $(1+r)^{-1}$. By introducing the dividend growth rate $g_t$, such that $D_t = (1+g_t)D_{t-1}$, this equation can be rewritten as:

$$\frac{P_t}{D_t} = \sum_{i=1}^{\infty}\frac{1}{(1+r)^i}E_t\left[\prod_{j=1}^{i}\left(1+g_{t+j}\right)\right]. \tag{2.2}$$

According to equation (2.2), any movements of the PD ratio $\left(\frac{P_t}{D_t}\right)$ can be caused only by time-variation of the discount factor or by changed expectations on future dividend growth rates. Under the assumption of a constant discount factor, an increase in the PD ratio should predict

an increase in future dividends and vice versa. However, Campbell and Shiller (2001) argue that neither the PD nor the PE ratio are good predictors for future dividend growth rates. Instead, both valuation ratios work well as a predictor for future stock returns. High valuation ratios predict decreasing stock prices, while low ratios predict increasing prices (Campbell and Shiller, 2001).

The assumption of a constant discount factor is very restrictive. Instead, modern asset pricing models often incorporate a stochastic discount factor (SDF), representing the time-varying risk aversion of a representative agent (Cochrane, 2011). Nevertheless, Campbell and Shiller (1988) show that the finding of excess volatility is robust to several time-varying discount factors, including discount factors based on consumption, output, interest rates and return volatility.

Brock and Hommes (1998) provide an alternative to the present-value relationship (2.1) and the SDF framework, by allowing asset prices to depend on the expectations of $H$ different types of boundedly rational agents:

$$P_t = \frac{1}{1+r}\sum_{h=1}^{H} G_{h,t} E_t^h \left[ P_{t+1} + D_{t+1} \right], \tag{2.3}$$

with $E_t^h [\cdot]$ representing the beliefs of agent type $h$. The fraction of agents following trading strategy $h$ at time $t$ is denoted by $G_{h,t}$. For analytical tractability, Brock and Hommes (1998) assume a constant discount factor. This model nests the standard present-value model; when all types have rational beliefs (i.e. $E_t^h [\cdot] = E_t [\cdot] \ \forall h$), model (2.3) reduces to (2.1). Boswijk et al. (2007) show that if dividends are specified as a geometric random walk process, model (2.3) can be reformulated as follows:

$$y_t = \frac{1}{1+r}\sum_{h=1}^{H} G_{h,t} E_t^h \left[ y_{t+1} \right], \tag{2.4}$$

in which $y_t$ is defined as the PD ratio in deviation from its fundamental value. The results of Campbell and Shiller (2001) suggest to estimate mispricings in the market as the PD ratio in

25

deviation from its long-run average:

$$y_t = \frac{P_t}{D_t} - \mu, \tag{2.5}$$

in which $\mu = \frac{1}{T} \sum_{t=1}^{T} \frac{P_t}{D_t}$ represents an estimate of the fundamental value of the PD ratio. $y_t$ gives the size of the bubble in the market, which can be negative as well as positive. The asset is over-valued if $y_t > 0$ and under-valued if $y_t < 0$. The price of the asset $P_t$ can be decomposed in an estimated fundamental value $\mu D_t$ and bubble $y_t D_t$:

$$P_t = \mu D_t + y_t D_t \tag{2.6}$$

A widely cited example of model (2.3) distinguishes two types of agents, fundamentalists and chartists, who are both aware of the fundamental value, but disagree about the persistence of the deviation from this fundamental value. The fundamentalists' strategy is to buy stocks when the market is undervalued and sell when the market is overvalued. They believe in mean reversion; mispricings in the market should disappear over time: $E_t^F[y_{t+1}] = \eta_F y_{t-1}$, with $\eta_F < 1 + r$. Chartists (or trend-followers), on the other hand, speculate that the stock market will continue to diverge from its fundamental valuation: $E_t^C[y_{t+1}] = \eta_C y_{t-1}$, with $\eta_C \geqslant 1 + r$.

By substituting these two beliefs into (2.4) and allowing the fractions of both agent types to vary over time, the asset pricing process can be described by a smooth-transition autoregressive (STAR) process:

$$y_t = \alpha_F y_{t-1}(1 - G_t) + \alpha_C y_{t-1} G_t + \varepsilon_t, \tag{2.7}$$

with $\alpha_F = \eta_F / (1 + r) < 1$ and $\alpha_C = \eta_C / (1 + r) \geqslant 1$. The transition function $G_t$ defines the fraction of chartist in the market. The fraction of fundamentalists is in this two-type model is given by $1 - G_t$. Although both types use a linear prediction rule, the time-varying fractions of each agent type makes the process nonlinear and, under certain parametrizations, chaotic (Brock and Hommes, 1998).

Boswijk et al. (2007) estimate a variant of this model for both the PD and PE ratio of the

S&P 500 index, in deviation from their mean, for the period 1871 to 2003. They follow Brock and Hommes (1998) by letting agents update their beliefs based on the realized profits of each type in the previous period. Under these evolutionary dynamics, agents switch from the less profitable strategy to the more profitable strategy. The transition function therefore becomes a logistic function depending on lagged values of the dependent variable:

$$G_t = (1 + exp[-\gamma(\eta_C - \eta_F)y_{t-3}(y_{t-1} - (1+r)y_{t-2})])^{-1}, \qquad (2.8)$$

in which $\gamma$ represents the intensity of choice of the agents. If $\gamma \to \infty$ all agents choose the strategy that was most profitable in the previous period. On the other hand, if $\gamma = 0$, the fraction of both types is exactly 50% in all periods, independent of the realized profits.

Instead of these evolutionary dynamics, I let the agents base their choice of strategy on macroeconomic and financial information, which can be interpreted as an extension of the agents' information set. Of interest is to find which economic conditions can be associated with each type of agent.

The transition function $G_t$ is a logistic function, as in the logistic STAR (LSTAR) model by Teräsvirta (1994):

$$G_t = (1 + exp[-\gamma(x_t - c)])^{-1}, \qquad (2.9)$$

in which the transition variable $x_t$ is usually a lagged value or lagged difference of the dependent variable, but can be any predetermined or exogenous variable. The transition function may also depend on a linear combination of variables:

$$G_t = (1 + exp[-\gamma(X_t\beta - c)])^{-1}, \qquad (2.10)$$

with $X_t = [x_{1,t} \ldots x_{p,t}]$ and $p$ is the number of included transition variables. For this model; $\gamma$, $c$ and $\beta$ can not be all identified. This problem can be solved by placing a restriction on $\beta$. In this paper, the elements of $\beta$ are restricted to sum to one, so that $X_t\beta$ is a weighted sum of multiple transition variables.
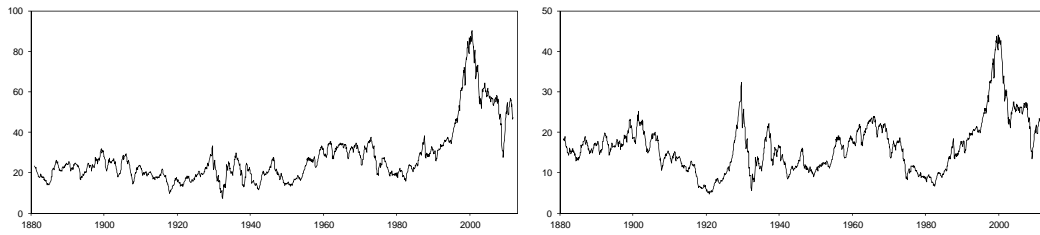
Figure 2.1: S&P 500 index 1881Q1-2011Q4: price-dividend ratio (left) and price-earnings ratio (right).

## 2.3 Data and linearity tests

Figure 2.1 shows quarterly data of the PD (left) and PE (right) ratios of the S&P500 index since 1881[2]. These valuation ratios show the level of the S&P500 index relative to the cash flows that the indexed stocks are generating. In particular the path of the PE ratio (right) seems stable or mean-reverting in the long run. Even after reaching record levels around the start of this century, the PE ratio recently dropped again below its average value during the credit crisis in 2009. This latest peak is comparable in size to earlier episodes, most notably the 1920s. For the PD ratio, this pattern is less clear. Due to relatively low dividend payouts by listed firms in recent decades (Fama and French, 2001), the PD ratio climbs during the 1990s to much higher levels than during any earlier peaks in the market. Although the model in Section 2.2 is expressed in terms of the PD ratio, I estimate the STAR model with both these valuation ratios as the dependent variable. Earnings are smoothed over a period of ten years, creating the so-called cyclically adjusted PE ratio. Both valuation ratios are taken in deviation from their average value.

I follow the specification, estimation and evaluation cycle for STAR models proposed by Teräsvirta (1994). The specification stage includes the selection of the appropriate lag structure and justification of STAR modeling by testing for linearity. To find the optimal lag length, I estimate linear AR($q$) models including up to six lags for both the PD and PE ratio. Table 2.1 shows the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) for all specifications. For both valuation ratios, the AR(1) model is selected as the appropriate specification. The STAR model is therefore estimated with an autoregressive structure of one

---

[2]Source: Robert Shiller, http://www.irrationalexuberance.com/index.htm

**TABLE 2.1:** AR(q): Selection criteria

| $y_t$ | q: | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $PD_t$ | AIC | -699.5 | -696.7 | -691.2 | -686.7 | -680.5 | -676.7 |
| | BIC | -692.8 | -686.7 | -677.8 | -670.1 | -660.6 | -653.5 |
| $PE_t$ | AIC | -681.8 | -678.1 | -672.4 | -669.7 | -664.9 | -662.1 |
| | BIC | -675.2 | -668.1 | -659.1 | -653.1 | -645.0 | -638.9 |

*Notes*: Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) for AR($q$) models. Sample size (for $y_t = PD_t$ and $y_t = PE_t$) is 208 observations: 1960Q1-2011Q4.

lag, as in equation (2.7). At the end of this paper, I verify the sufficiency of this lag structure by submitting the residuals from the final STAR model to a test of serial independence.

The next step is to test for linearity and simultaneously select the transition variables. I consider a set of financial and macroeconomic indicators as potential transition variables[3]. The first set of indicators is related to the performance of the stock market and includes both dependent variables (*PD* and *PE*), monthly returns (*RET*) and the volatility of the S&P500 index (*VOL*), defined as the variance of daily returns in each quarter. For the other indicators I follow the choice of variables by Campbell (2003), who uses business cycle indicators, inflation and interest rates to study the cyclical properties of risk premiums. The business cycle indicators considered by Campbell (2003) are real GDP (*GDP*) and consumption (*CON*). I supplement these indicators with the output gap (*OPG*) and industrial production (*IND*). The inflation rates are the consumer price index (*CPI*) and GDP deflator (*DEF*). The interest rates used by Campbell (2003) are the short-term yield on 3-month US treasury bills (*STY*) and the long-term yield on 10-year US treasury notes (*LTY*). I add to this the 10-year yield on Baa-rated corporate bonds (*CBY*) and construct the term spread ($TSP = LTY - STY$) and the yield spread of corporate bonds over sovereign bonds ($YSP = CBY - LTY$). While the business cycle indicators measure the current state of the economy, these interest rates and spreads contain expectations on future macroeconomic conditions (Bernanke, 1990; Estrella and Mishkin, 1998). *GDP*, *CON*, *IND*, *CPI* and *DEF* are measured in quarter-on-quarter growth rates. *OPG* is a percentage of GDP. For the interest rates and the output gap I look at both levels and first differences (denoted by $\triangle$).

---

[3] Source: FRED® (Federal Reserve Economic Data)

These data are not available for the full period of S&P500 data, so the model is estimated using 208 observations (1960Q1-2011Q4). All variables are standardized (demeaned and divided by their standard deviation), to accommodate numerical estimation of the nonlinear model. For all explanatory variables, I consider both first and second lags, which are therefore predetermined with respect to the dependent variable.

To determine which of these variables are valid transition variables in the STAR model, they are submitted to a linearity test based on a Taylor approximation of the STAR model following Luukkonen et al. (1988). First, I consider the univariate transition function (2.9). A third-order Taylor approximation of (2.7) with univariate transition function (2.9) around $\gamma = 0$ gives:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \sum_{i=1}^{3} \phi_{1+i} y_{t-1} x_t^i + e_t. \tag{2.11}$$

Linearity can now be tested by estimating this Taylor approximation by OLS and testing the null hypothesis $H_o : \phi_2 = \phi_3 = \phi_4 = 0$. Rejection of linearity implies that $x_t$ is a valid transition variable.

Results of the linearity tests are given in Table 2.2, which shows the test statistics and corresponding P-values. The test statistic is asymptotically $F(n, T - k - n - 1)$ distributed under the null, with $T = 208$ (observations), $k = 2$ (unrestricted parameters) and $n = 3$ (restricted parameters). An asymptotically equivalent $\chi^2$-test may be applied here as well, but the F-test has preferable properties in small samples (Teräsvirta et al., 2010). The results in Table 2.2 show that several variables are valid transition variables.

I consider the LSTAR only, since a logistic transition function follows directly from the logit switching rule in the model by Brock and Hommes (1998). Alternatively, the transition function could be an exponential function as in the ESTAR model. To verify that the LSTAR is the correct model, I apply a sequence of three F-tests based on (2.11) proposed by Teräsvirta (1994) to choose between both transition functions: $H_{o1} : \phi_4 = 0$, $H_{o2} : \phi_3 = 0 \mid \phi_4 = 0$ and $H_{o4} : \phi_2 = 0 \mid \phi_3 = \phi_4 = 0$. If $H_{02}$ yields a stronger rejection than $H_{01}$ and $H_{03}$, the ESTAR model is the best choice. Otherwise, the LSTAR model is preferred. Table 2.2 shows that with

TABLE 2.2: Linearity tests: Univariate transition function

| | $y_t = PD_t$ | | | | | | $y_t = PE_t$ | | | | | |
| lag | $t-1$ | | | $t-2$ | | | $t-1$ | | | $t-2$ | | |
| $x$ | F | P | L/E | F | P | L/E | F | P | L/E | F | P | L/E |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PD | 0.667 | 0.573 | L | 0.974 | 0.406 | L | 3.359 | 0.020 | E | 2.811 | 0.041 | E |
| PE | 0.236 | 0.871 | E | 0.282 | 0.838 | L | 0.512 | 0.674 | L | 0.475 | 0.700 | L |
| RET | 2.407 | 0.068 | E | 0.600 | 0.616 | L | 2.741 | 0.044 | E | 0.266 | 0.850 | E |
| VOL | 1.621 | 0.186 | L | 0.818 | 0.486 | L | 0.496 | 0.686 | L | 0.541 | 0.655 | L |
| GDP | 4.742 | 0.003 | L | 0.868 | 0.459 | L | 3.495 | 0.017 | L | 0.574 | 0.633 | L |
| CON | 2.596 | 0.054 | L | 0.873 | 0.456 | L | 0.849 | 0.469 | L | 0.484 | 0.694 | E |
| OPG | 1.555 | 0.202 | L | 0.337 | 0.799 | L | 0.483 | 0.694 | E | 1.820 | 0.145 | E |
| $\triangle$OPG | 3.847 | 0.010 | L | 0.760 | 0.518 | L | 3.299 | 0.021 | L | 0.614 | 0.607 | L |
| IND | 5.073 | 0.002 | L | 2.845 | 0.039 | L | 4.358 | 0.005 | L | 2.249 | 0.084 | L |
| CPI | 1.119 | 0.342 | L | 1.084 | 0.357 | L | 1.261 | 0.289 | L | 0.732 | 0.534 | L |
| DEF | 2.639 | 0.051 | L | 1.201 | 0.311 | L | 4.102 | 0.007 | L | 1.472 | 0.223 | L |
| STY | 1.139 | 0.334 | L | 1.247 | 0.294 | L | 1.205 | 0.309 | L | 1.339 | 0.263 | L |
| $\triangle$STY | 0.254 | 0.858 | L | 1.475 | 0.223 | L | 0.162 | 0.922 | L | 0.577 | 0.631 | L |
| LTY | 0.238 | 0.870 | L | 0.577 | 0.631 | E | 0.283 | 0.838 | L | 0.833 | 0.477 | L |
| $\triangle$LTY | 0.496 | 0.686 | L | 0.565 | 0.639 | L | 0.335 | 0.800 | L | 0.519 | 0.670 | L |
| TSP | 2.591 | 0.054 | L | 2.724 | 0.045 | L | 1.476 | 0.222 | E | 1.498 | 0.216 | L |
| CBY | 0.128 | 0.943 | E | 0.163 | 0.921 | E | 0.056 | 0.982 | L | 0.071 | 0.975 | E |
| $\triangle$CBY | 0.391 | 0.760 | L | 0.076 | 0.973 | L | 0.109 | 0.955 | L | 0.354 | 0.787 | L |
| YSP | 1.414 | 0.240 | L | 1.971 | 0.119 | L | 1.375 | 0.252 | L | 2.216 | 0.087 | L |

*Notes*: F-test statistics and corresponding P-values for $H_o : \phi_2 = \phi_3 = \phi_4 = 0$ in equation (2.11), using both first and second lags of several transition variables. L/E refers to the LSTAR or ESTAR model selected by the procedure of Teräsvirta (1994).

most transition variables, the LSTAR (marked by L) is the preferred specification. Teräsvirta (1994) further recommends to estimate the STAR model with the transition variable for which rejection of linearity is the strongest. However, the fact that linearity is rejected for different transition variables suggests to incorporate more than one variable in the transition function.

Allowing for a multivariate transition function, I now propose a similar procedure based on linearity tests to select the appropriate transition variables $X = [x_1 \ldots x_p]$. From substituting $x_t = X_t \beta$ into (2.11) it becomes clear that this Taylor approximation can not be estimated by OLS if the weights $\beta$ are unknown. To circumvent this problem, I first estimate $\beta$ based on a first-order Taylor approximation[4] of (2.7), with a multivariate transition function (2.10) around

---

[4]A linearity test based on a first-order Taylor approximation does not allow to choose between a LSTAR and ESTAR, but does provide power against STAR nonlinearity in general, except when the regime switching is in the intercept rather than the autoregressive parameters (Luukkonen et al., 1988).

$\gamma = 0$:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-1}(X_t \beta) + e_t, \qquad (2.12)$$

or:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \sum_{i=1}^{p} \theta_i y_{t-1} x_{i,t-1} + e_t, \qquad (2.13)$$

such that $\theta_i = \phi_2 \beta_i$. This Taylor approximation can be estimated by OLS for any set of explanatory variables, after which the OLS estimates $\widehat{\theta}$ and the restriction $\sum_{i=1}^{p} \beta_i = 1$ can be used to derive estimates of $\beta$:

$$\theta_i = \phi_2 \beta_i$$

$$\sum_{i=1}^{p} \theta_i = \phi_2 \sum_{i=1}^{p} \beta_i = \phi_2$$

$$\widetilde{\beta}_j = \left( \sum_{i=1}^{p} \widehat{\theta}_i \right)^{-1} \widehat{\theta}_j. \qquad (2.14)$$

Selecting the optimal set of transition variables consists of the following steps. First, I estimate (2.13) for each possible set of one to four transition variables, which never includes more than one variable out of each of the following four groups: (*i*) Stock market indicators, (*ii*) business cycle indicators, (*iii*) inflation rates and (*iv*) interest rates and spreads. This approach limits the number of sets under consideration and, because several variables within each group are highly correlated, it avoids multicollinarity within the transition function. For each set, I then compute $\widetilde{\beta}$, following (2.14) and perform a t-test on each element of $\widetilde{\beta}$. In trying to avoid selecting an overfitted model, I proceed only with those sets of variables for which all elements of $\widetilde{\beta}$ are significant at the 10% level. For these selected sets, I substitute $x_t = X_t \widetilde{\beta}$ into the third-order Taylor approximation (2.11) in order to test the null hypothesis $H_o : \phi_2 = \phi_3 = \phi_4 = 0$. Finally, I choose the set of variables yielding the strongest rejection of linearity as the optimal set of transition variables. Table 2.3 reports the final results of this test procedure. With the selected linear combinations of transition variables, the rejection of linearity is stronger than with any of the single transition variables in Table 2.2. In both cases the LSTAR model is preferred over the ESTAR.

**TABLE 2.3:** Linearity tests: Multivariate transition function

| $y_t$ | $X_t$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $F$ | $P$ | L/E |
|---|---|---|---|---|---|---|---|
| $PD_t$ | $(VOL_{t-1}, IND_{t-1}, STY_{t-2})$ | 0.20 | 0.54 | 0.26 | 7.98 | $4.7 \times 10^{-5}$ | L |
| $PE_t$ | $(IND_{t-1}, DEF_{t-2})$ | 0.67 | 0.33 | . | 7.79 | $6.0 \times 10^{-5}$ | L |

*Notes*: Optimal set of transition variables $X_t$ in terms of the highest F-test statistics and lowest P-values for $H_o$ : $\phi_2 = \phi_3 = \phi_4 = 0$ in equation (2.11), with $x_t = X_t \beta$. L/E refers to the LSTAR or ESTAR model selected by the procedure of Teräsvirta (1994). The elements of $\beta$ are estimated based on equations (2.13)-(2.14)

## 2.4 Results

The parameter estimates for the STAR model are presented in Table 2.4. The models are estimated by nonlinear least squares, preceded by a $(p + 1)$-dimensional grid search for $\gamma$, $c$ and the $(p - 1)$ free elements of $\beta$ to find starting values. The selection criterion in this grid search is the sum of squares of the STAR model, which can be estimated by OLS when $\gamma$, $c$ and $\beta$ are kept fixed. The estimated autoregressive parameters of each regime are denoted by $\alpha_1$ and $\alpha_2$, rather than $\alpha_C$ and $\alpha_F$, because the latter notation implies restrictions on these parameters that I do not impose during estimation.

**TABLE 2.4:** Parameter estimates for STAR model

| $y_t$ | $X_t$ | $\alpha_1$ | $\alpha_2$ | $\gamma$ | $c$ | $\beta_1$ | $\beta_2$ | $\beta_3$ |
|---|---|---|---|---|---|---|---|---|
| $PD_t$ | $IND_{t-1}$ | 0.948 | 1.098 | 80.44 | 0.375 | . | . | . |
|  |  | (0.010) | (0.021) | (52.79) | (0.012) | . | . | . |
| $PE_t$ | $IND_{t-1}$ | 0.898 | 1.019 | 1244 | -0.371 | . | . | . |
|  |  | (0.016) | (0.011) | (1247) | (2.148) | . | . | . |
| $PD_t$ | $(VOL_{t-1}, IND_{t-1}, STY_{t-2})$ | 0.917 | 1.101 | 7.452 | 0.123 | -0.012 | 0.721 | 0.291 |
|  |  | (0.017) | (0.026) | (2.572) | (0.089) | (0.076) | (0.077) | (0.040) |
| $PE_t$ | $(IND_{t-1}, DEF_{t-2})$ | 0.841 | 1.045 | 4.739 | -0.372 | 0.656 | 0.344 | . |
|  |  | (0.036) | (0.023) | (1.873) | (0.135) | (0.069) | (0.069) | . |

*Notes*: NLS parameter estimates for model (2.7) with univariate transition function (2.9) or multivariate transition function (2.10). Standard errors in parenthesis. All estimated models include a constant, which are not significantly different from zero and are therefore not reported.

The top rows of Table 2.4 show the parameter estimates for the STAR models (2.7) with univariate transition function (2.9), using the transition variable for which rejection of linearity is the strongest, which is the first lag of industrial production ($IND_{t-1}$) for both valuation ratios. Because there is only one transition variable, there are no weights $\beta$ to estimate. Although both estimated models include a mean-reverting and a trend-following regime, the results are not

entirely consistent with the spirit of the heterogeneous agent model by Brock and Hommes (1998), because the intensity of choice parameter $\gamma$ is so high that the fraction of each type is either zero or one. Contrary to the idea of heterogeneous beliefs these results suggest that the entire population of agents makes the same switch simultaneously.

The bottom rows of Table 2.4 show the STAR models (2.7) with multivariate transition function (2.10). With multiple transition variables, the estimates of $\gamma$ are lower, in support of a smooth transition between the regimes. In both estimated models, two distinct regimes are identified. Each specification has one autoregressive parameter significantly smaller than one (representing the fundamentalist type), while the other autoregressive parameter is significantly greater than one (representing the chartist type). Interpreting $\beta$ reveals that chartists are more dominant during periods of economic expansion, while the fraction of fundamentalists increases during economic downturns.

With $y_t = PD_t$, the effect of volatility ($VOL_{t-1}$) does not seem significant. I keep this transition variable in the model, because excluding it does not improve the fit of the model. Industrial production growth ($IND_{t-1}$) has a positive coefficient, implying in this case it supports the chartist type. An increase in industrial production causes an increase in the fraction of chartists in the economy. Also the short-term yield on 3-month treasury bills ($STY_{t-2}$) has a positive coefficient. A high yield on low-risk assets like treasury bills implies low levels of risk aversion, and in this model a high fraction of chartists. With $y_t = PE_t$, the model does not include the exact same set of transition variables, but the results tell a similar story: Chartism is the dominant strategy during expansive periods, signaled by high industrial production growth ($IND_{t-1}$) and inflation ($DEF_{t-2}$).

Several measures are applied to evaluate the fit of the STAR model, compared to the fit of an AR(1) model and the linear regression model:

$$y_t = \omega_1 y_{t-1} + X_t \omega_2 + e_t, \tag{2.15}$$

which includes the same explanatory variables as the STAR model. Table 2.5 presents, in

34

addition to the $R^2$, AIC and BIC of all models, the results of a pseudo out-of-sample forecasting exercise. Using an expanding window approach, I estimate all models using a subset of the data (1960Q2-$S$) and use the estimated models to compute forecasts for period $S+1$. This process is repeated 48 times, creating pseudo out-of-sample forecasts for the period (2000Q1-2011Q4), from which Mean Absolute Errors (MAE) and Root Mean Squared Errors (RMSE) are computed. Due to the high persistency of the valuation ratios, the $R^2$ of all models including the univariate AR(1) are relatively high. The improved fit of the STAR model over the linear alternatives is small but seems robust to several measures. According to the AIC, BIC and out-of-sample results, the STAR model with multivariate transition function outperforms its linear alternatives as well as the STAR model with a univariate transition function. The result that the STAR model (2.7)-(2.10) has a better fit than the linear model (2.15) implies that the variables in $X_t$ work better in explaining the switching process between mean-reverting and trend-following regimes than they do in explaining the level of $PD_t$ and $PE_t$, which supports the notion of chartism and fundamentalism. The macroeconomic information is not simply correlated with stock prices but has an effect on the nonlinear adjustment towards the fundamental value. Table 2.5 also shows the test statistics and bootstrap P-values for the linearity test by Hansen (1996, 1997). Like the linearity tests in Section 2.2, these tests show strong rejections of linearity, with P-values lower than 1%.

An intuitive interpretation of the results is found by giving (2.7) the alternative formulation of an AR(1) process with a time-varying parameter:

$$y_t = \delta_t y_{t-1} + \varepsilon_t, \tag{2.16}$$

in which $\delta_t = \alpha_1(1 - G_t) + \alpha_2 G_t$, which can be interpreted as an indicator of market sentiment. When $\delta_t \geqslant 1$ the valuation ratio is diverging from its mean, implying that the chartist regime is dominant, while the valuation ratio is mean-reverting when $\delta_t < 1$. Figure 2.2 offers a graphical evaluation of both estimated models by showing plots of $\delta_t$ over time and scatter plots of $G_t$ against $X'_{t-1}\beta$, evaluated at the estimates of the multivariate STAR model. Because of the

**TABLE 2.5:** Goodness of fit

| $y_t$ | $X_t$ | model | $R^2$ | AIC | BIC | MAE | RMSE | $F^{lin}$ | $P$ (boot) |
|-------|-------|-------|-------|-----|-----|-----|------|-----------|------------|
| $PD_t$ | . | AR(1) | 0.966 | -699.5 | -692.8 | 1.317 | 1.526 | . | . |
| $PD_t$ | $IND_{t-1}$ | Linear | 0.966 | -697.5 | -687.5 | 1.321 | 1.532 | . | . |
| $PD_t$ | $IND_{t-1}$ | STAR | 0.970 | -718.0 | -704.7 | 1.292 | 1.490 | 23.81 | 0.002 |
| $PD_t$ | $(VOL_{t-1}, IND_{t-1}, STY_{t-2})$ | Linear | 0.967 | -699.1 | -682.4 | 1.323 | 1.538 | . | . |
| $PD_t$ | $(VOL_{t-1}, IND_{t-1}, STY_{t-2})$ | STAR | 0.971 | -723.3 | -710.0 | 1.283 | 1.490 | 29.79 | 0.001 |
| | | | | | | | | | |
| $PE_t$ | . | AR(1) | 0.963 | -681.8 | -675.2 | 0.943 | 1.227 | . | . |
| $PE_t$ | $IND_{t-1}$ | Linear | 0.963 | -679.9 | -669.9 | 0.946 | 1.231 | . | . |
| $PE_t$ | $IND_{t-1}$ | STAR | 0.966 | -696.1 | -682.7 | 0.919 | 1.196 | 19.06 | 0.003 |
| $PE_t$ | $(IND_{t-1}, DEF_{t-2})$ | Linear | 0.965 | -686.1 | -672.8 | 0.940 | 1.216 | . | . |
| $PE_t$ | $(IND_{t-1}, DEF_{t-2})$ | STAR | 0.967 | -701.1 | -687.8 | 0.904 | 1.193 | 24.62 | 0.002 |

*Notes*: Measures of goodness of fit of the STAR models from Table 2.4, a linear AR(1) model and the linear models (2.15) including the same explanatory variables as the STAR. Mean Absolute Errors and Root Mean Square Errors are computed from 48 pseudo out-of-sample forecasts for 2000Q1-2011Q4. The F-test for linearity by Hansen (1996, 1997) tests $H_o : \alpha_1 = \alpha_2$ in the STAR model. The corresponding bootstrap P-value is computed based on 10.000 replications.

relatively low value of the intensity of choice parameter $\gamma$, both scatter plots on the right side of Figure 2.2 clearly show a logistic curve. Most of the time, both chartists and fundamentalists are represented in the economy, with $\delta_t$ fluctuating around one. In 2001 and again in 2008 the market turned almost completely to the fundamentalist type for a prolonged period, causing the bubble built up in the 1990s to deflate.

Finally, the estimated multivariate models in Table 2.4 are evaluated with diagnostic checks. Table 2.6 presents results on tests of serial independence, parameter constancy and no remaining nonlinearity. Eitrheim and Teräsvirta (1996) provide technical details on all three tests.

The test of serial independence test the null hypothesis of no $q^{th}$ order autocorrelation in the residuals. For a $q^{th}$ order test, the resulting test statistic is asymptotically $F(q, T - q - 4)$ distributed under the null, with $T = 208$ (sample size). I execute this test for first- up to fourth-order autocorrelation. For both models, the test results give no reason the reject the null hypothesis, confirming the sufficiency of an autoregressive structure of only one lag.

Under the null hypothesis of no time-variation of the parameters in (2.7) and (2.10), the parameter constancy test statistic is asymptotically $F(6, T - 10)$ distributed. Also this test gives no reason to reject the specification.
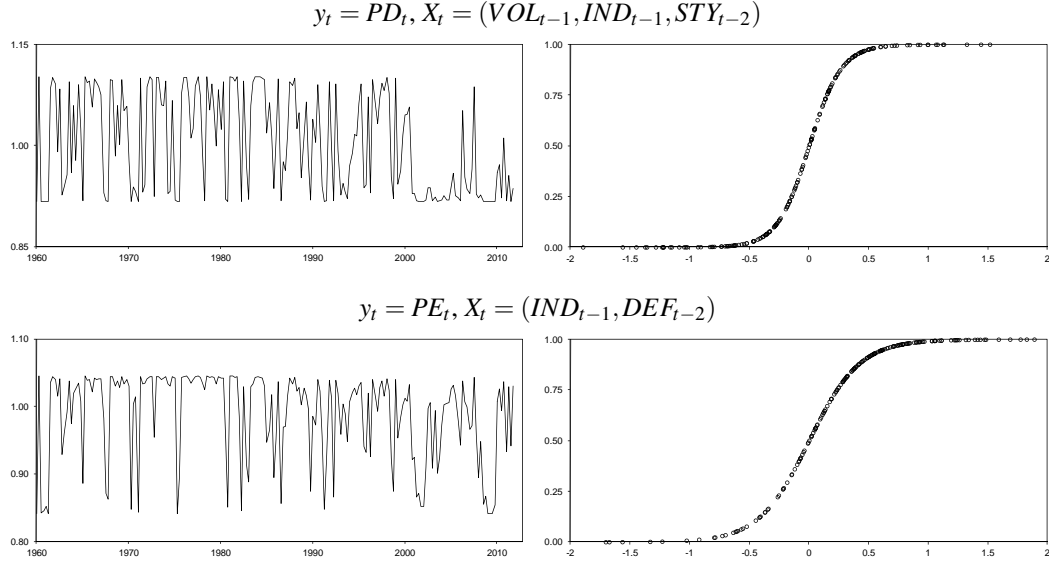
$y_t = PD_t, X_t = (VOL_{t-1}, IND_{t-1}, STY_{t-2})$

$y_t = PE_t, X_t = (IND_{t-1}, DEF_{t-2})$

Figure 2.2 :Regression results: Plot (left) of $\delta_t = \alpha_1(1 - G_t) + \alpha_2 G_t$ over time and scatterplot (right) of $G_t$ against $X_t\beta$, evaluated at parameter estimates in Table 2.4.

The test of no remaining nonlinearity checks whether any variable has a significant nonlinear effect on the residuals. This could be the case when a transition variable is omitted, or when these variables have an effect on $y_t$ through some other nonlinear channel. The test statistic is asymptotically $F(3, T - 6)$ distributed under the null. This test is repeated for the first lags of all potential transition variables considered in this paper. For the majority of the variables, the null hypothesis of no remaining non-linearity can not be rejected at the 10% level. There are some exceptions, in particular lagged returns ($RET_{t-1}$), but including these variables in the transition function does not improve the fit of the model. Given that the test is repeated for many variables, it is possible that the rejections are Type I errors. Overall, the results of these diagnostic checks are positive and provide support to the specification of the model.

## 2.5   Conclusion

In this paper, I identify two types of agents: fundamentalists and chartists. The presence of chartists, who are predicting trends rather than fundamentals, explains the existence of bubbles in asset prices. To estimate the effects of macroeconomic conditions on the behavior of agents, I propose a STAR model with a multivariate transition function. This STAR model outperforms

**TABLE 2.6:** Diagnostic tests

| | $y_t$ | $PD_t$ | | $PE_t$ | |
|---|---|---|---|---|---|
| | $X_t$ | $(VOL_{t-1}, IND_{t-1}, STY_{t-2})$ | | $(IND_{t-1}, DEF_{t-2})$ | |
| | | F | P | F | P |
| Serial independence: | $1^{st}$ | 1.380 | 0.242 | 1.327 | 0.251 |
| | $2^{nd}$ | 0.804 | 0.449 | 0.805 | 0.448 |
| | $3^{rd}$ | 0.921 | 0.432 | 1.683 | 0.172 |
| | $4^{th}$ | 0.846 | 0.498 | 1.250 | 0.291 |
| Parameter constancy: | | 1.225 | 0.295 | 1.529 | 0.170 |
| No remaining nonlinearity: | $PD_{t-1}$ | 1.210 | 0.307 | 4.195 | 0.007 |
| | $PE_{t-1}$ | 0.389 | 0.761 | 2.974 | 0.033 |
| | $RET_{t-1}$ | 4.878 | 0.003 | 4.816 | 0.003 |
| | $VOL_{t-1}$ | 2.267 | 0.082 | 0.651 | 0.583 |
| | $GDP_{t-1}$ | 0.835 | 0.476 | 0.943 | 0.421 |
| | $CON_{t-1}$ | 0.639 | 0.591 | 0.326 | 0.807 |
| | $OPG_{t-1}$ | 0.425 | 0.735 | 0.445 | 0.721 |
| | $\triangle OPG_{t-1}$ | 0.635 | 0.593 | 0.837 | 0.475 |
| | $IND_{t-1}$ | 0.126 | 0.945 | 0.645 | 0.587 |
| | $CPI_{t-1}$ | 1.231 | 0.299 | 1.478 | 0.222 |
| | $DEF_{t-1}$ | 2.131 | 0.097 | 4.832 | 0.003 |
| | $STY_{t-1}$ | 0.090 | 0.966 | 0.616 | 0.605 |
| | $\triangle STY_{t-1}$ | 0.277 | 0.842 | 1.730 | 0.162 |
| | $LTY_{t-1}$ | 0.778 | 0.508 | 0.459 | 0.711 |
| | $\triangle LTY_{t-1}$ | 0.200 | 0.896 | 0.886 | 0.449 |
| | $TSP_{t-1}$ | 1.192 | 0.314 | 1.283 | 0.281 |
| | $CBY_{t-1}$ | 0.811 | 0.489 | 0.472 | 0.702 |
| | $\triangle CBY_{t-1}$ | 0.577 | 0.631 | 0.164 | 0.920 |
| | $YSP_{t-1}$ | 0.469 | 0.704 | 0.048 | 0.986 |

*Notes*: F-test statistics and corresponding P-values for first- to fourth-order serial independence, parameter constancy and no remaining non-linearity (Eitrheim and Teräsvirta, 1996)

STAR models with a single transition variable as well as linear alternatives in terms of goodness-of-fit.

Agents are more willing to believe in the persistence of bubbles during times of positive macroeconomic news. Chartists gain dominance during periods of favorable economic conditions, mainly measured by industrial production. The fraction of fundamentalists increases during economic downturns, which encourage agents to re-appreciate fundamentals.

Further research in this area may include an investigation of international stock markets, in order to find whether the switching between chartism and fundamentalism is based on the same

factors and occurs simultaneously across countries. In addition, the framework presented in this paper is suitable to find the macroeconomic conditions under which any asset price deviates from some measure of fundamental value. Other possible applications include the deviation of exchange rates from purchasing power parity (see e.g. Rogoff, 1996), or the term structure of interest rates in deviation from the expectations hypothesis (see e.g. Mankiw and Miron, 1986).

# References

Becker, R. and D. Osborn: 2012, 'Weighted smooth transition regressions'. *Journal of Applied Econometrics*.

Bernanke, B. S.: 1990, 'On the predictive power of interest rates and interest rate spreads'. *New England Economic Review* (Nov), 51–68.

Boswijk, H. P., C. H. Hommes, and S. Manzan: 2007, 'Behavioral heterogeneity in stock prices'. *Journal of Economic Dynamics and Control* **31**(6), 1938–1970.

Brock, W. A. and C. H. Hommes: 1997, 'A Rational Route to Randomness'. *Econometrica* **65**(5), 1059–1096.

Brock, W. A. and C. H. Hommes: 1998, 'Heterogeneous beliefs and routes to chaos in a simple asset pricing model'. *Journal of Economic Dynamics and Control* **22**(8-9), 1235–1274.

Campbell, J.: 2003, 'Consumption-based asset pricing'. In: G. Constantinides, M. Harris, and R. Stulz (eds.): *Handbook of the Economics of Finance*, Vol. 1B. Elsevier, Chapt. 13.

Campbell, J. Y. and R. J. Shiller: 1988, 'The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors'. *Review of Financial Studies* **1**(3), 195–228.

Campbell, J. Y. and R. J. Shiller: 2001, 'Valuation Ratios and the Long-Run Stock Market Outlook: An Update'. *NBER Working Papers* (8221).

Chiarella, C. and X. He: 2002, 'Heterogeneous beliefs, risk and learning in a simple asset pricing model'. *Computational Economics* **19**(1), 95–132.

Chiarella, C., G. Iori, and J. Perelló: 2009, 'The impact of heterogeneous trading rules on the limit order book and order flows'. *Journal of Economic Dynamics and Control* **33**(3), 525–537.

Chordia, T. and L. Shivakumar: 2002, 'Momentum, business cycle, and time-varying expected returns'. *The Journal of Finance* **57**(2), 985–1019.

Cochrane, J. H.: 2011, 'Presidential Address: Discount Rates'. *The Journal of Finance* **66**(4), 1047–1108.

Cooper, I. and R. Priestley: 2009, 'Time-varying risk premiums and the output gap'. *Review of Financial Studies* **22**(7), 2801–2833.

De Jong, E., W. F. Verschoor, and R. C. Zwinkels: 2010, 'Heterogeneity of agents and exchange rate dynamics: Evidence from the EMS'. *Journal of International Money and Finance* **29**(8), 1652–1669.

Eitrheim, O. and T. Teräsvirta: 1996, 'Testing the adequacy of smooth transition autoregressive models'. *Journal of Econometrics* **74**(1), 59–75.

Estrella, A. and F. S. Mishkin: 1998, 'Predicting U.S. Recessions: Financial Variables As Leading Indicators'. *The Review of Economics and Statistics* **80**(1), 45–61.

Fama, E. and K. French: 1989, 'Business conditions and expected returns on stocks and bonds'. *Journal of Financial Economics* **25**(1), 23–49.

Fama, E. F. and K. R. French: 2001, 'Disappearing dividends: changing firm characteristics or lower propensity to pay?'. *Journal of Financial Economics* **60**(1), 3–43.

Frijns, B., T. Lehnert, and R. C. Zwinkels: 2010, 'Behavioral heterogeneity in the option market'. *Journal of Economic Dynamics and Control* **34**(11), 2273–2287.

Gordon, M. J.: 1959, 'Dividends, Earnings, and Stock Prices'. *The Review of Economics and Statistics* **41**(2), 99–105.

Hansen, B.: 1996, 'Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis'. *Econometrica* **64**(2), 413–430.

Hansen, B.: 1997, 'Inference in TAR models'. *Studies in nonlinear dynamics and econometrics* **2**(1), 1–14.

Hommes, C. H.: 2006, 'Heterogeneous Agent Models in Economics and Finance'. In: L. Tesfatsion and K. L. Judd (eds.): *Handbook of Computational Economics*, Vol. 2. Elsevier, Chapt. 23.

LeRoy, S. F. and R. D. Porter: 1981, 'The Present-Value Relation: Tests Based on Implied Variance Bounds'. *Econometrica* **49**(3), 555–74.

Lof, M.: 2012, 'Heterogeneity in stock prices: A STAR model with multivariate transition function'. *Journal of Economic Dynamics and Control* **36**(12), 1845 – 1854.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta: 1988, 'Testing linearity against smooth transition autoregressive models'. *Biometrika* **75**(3), 491–499.

Mankiw, N. G. and J. A. Miron: 1986, 'The Changing Behavior of the Term Structure of Interest Rates'. *The Quarterly Journal of Economics* **101**(2), 211–28.

Manzan, S.: 2009, 'Agent Based Modeling in Finance'. In: R. A. Meyers (ed.): *Encyclopedia of Complexity and Systems Science*. Springer New York, pp. 3374–3388.

Manzan, S. and F. H. Westerhoff: 2007, 'Heterogeneous expectations, exchange rate dynamics and predictability'. *Journal of Economic Behavior & Organization* **64**(1), 111–128.

Medeiros, M. and A. Veiga: 2005, 'A flexible coefficient smooth transition time series model'. *Neural Networks, IEEE Transactions on* **16**(1), 97 –113.

Reitz, S., J. C. Rulke, and M. P. Taylor: 2011, 'On the Nonlinear Influence of Reserve Bank of Australia Interventions on Exchange Rates'. *The Economic Record* **87**(278), 465–479.

Reitz, S. and U. Slopek: 2009, 'Non-Linear Oil Price Dynamics: A Tale of Heterogeneous Speculators?'. *German Economic Review* **10**, 270–283.

Reitz, S. and M. Taylor: 2008, 'The coordination channel of foreign exchange intervention: a nonlinear microstructural analysis'. *European Economic Review* **52**(1), 55–76.

Reitz, S. and F. Westerhoff: 2003, 'Nonlinearities and cyclical behavior: the role of chartists and fundamentalists'. *Studies in Nonlinear Dynamics and Econometrics* **7**(4), 3.

Reitz, S. and F. Westerhoff: 2007, 'Commodity price cycles and heterogeneous speculators: a STAR-GARCH model'. *Empirical Economics* **33**(2), 231–244.

Rogoff, K.: 1996, 'The Purchasing Power Parity Puzzle'. *Journal of Economic Literature* **34**(2), 647–668.

Shiller, R. J.: 1981, 'Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?'. *American Economic Review* **71**(3), 421–36.

Spierdijk, L., J. A. Bikker, and P. van den Hoek: 2012, 'Mean reversion in international stock markets: An empirical analysis of the 20th century'. *Journal of International Money and Finance* **31**(2), 228 – 249.

Ter Ellen, S. and R. Zwinkels: 2010, 'Oil price dynamics: A behavioral finance approach with heterogeneous agents'. *Energy Economics* **32**(6), 1427–1434.

Teräsvirta, T.: 1994, 'Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models'. *Journal of the American Statistical Association* **89**(425), 208–218.

Teräsvirta, T., D. Tjostheim, and C. W. J. Granger: 2010, *Modelling Nonlinear Economic Time Series*. Oxford University Press.

West, K. D.: 1988, 'Dividend Innovations and Stock Price Volatility'. *Econometrica* **56**(1), 37–61.

# Chapter 3

# Rational speculators, contrarians and excess volatility[1]

## 3.1   Introduction

Prices of financial assets are typically more volatile than real economic activity. As a result, it is often impossible to associate asset price fluctuations with news regarding dividends underlying the asset. This excess volatility of asset prices with respect to dividends has been documented in many studies, such as Shiller (1981), Campbell and Shiller (1987), West (1988), or the survey by Gilles and LeRoy (1991). The behavioral finance literature has proposed various models to accommodate this excess volatility as well as other market anomalies (See e.g. the surveys by Hirshleifer, 2001, Barberis and Thaler, 2003, and Shiller, 2003). In such models, price movements can occur due to investor sentiment rather than fundamental news. Agents may make investment decisions based on expected price movements in the short run rather than expected dividends in the long run and often form non-rational expectations based on limited information sets and underparameterized models (See e.g. De Long et al., 1990a,b, Barberis et al., 1998, or Hong and Stein, 1999).

I consider a simple asset pricing model with three types of agents: Rational long-term investors, rational speculators and contrarians. These agents are allowed to have heterogeneous investment horizons and may form heterogeneous expectations regarding short-term price

---

[1]This chapter is based on *HECER Discussion Paper* 358 (Lof, 2012b)

movements. Nevertheless, all three types hold identical information sets and have their expectation formation mechanisms anchored in the same vector autoregressive (VAR) representation of prices and dividends. The model can therefore be evaluated empirically using the VAR approach for testing present value models, pioneered by Campbell and Shiller (1987, 1988), for which I use a dataset containing annual observations on the S&P500 index and underlying dividends for the period 1871-2011[2]. Even if there is no disagreement at all among the agents regarding expected dividends, the model is able to generate prices far more volatile than the standard present value model. Statistical tests indicate that the model is preferred to alternative representative agent models in which only one of the considered expectation formation mechanisms exists.

The first two agent types both act in accordance with the standard present value model. The only characteristic separating these agents is their investment horizon. The first type makes long-term investments and therefore values assets according to the cash flows (dividends) that the asset is expected to generate. I refer to agents of this type as rational long-term investors, while the term fundamentalism is also used in the literature to describe this behavior.

The second type is only interested in one-period returns, so that the main determinant of the asset's current value is the expected selling price in the next period. This speculative behavior is similar to that of the trend followers or the momentum traders considered in the literature, for example by Brock and Hommes (1998), or Hong and Stein (1999). However, while trend followers and momentum traders in general form expectations based on a simple univariate model and a limited information set, typically by extrapolating recent returns, the short-term investors considered in this paper form expectations by using the exact same model and information set as the rational long-term investors. I therefore refer to these agents as rational speculators.

I refer to the first two types of agents as rational, even if they are, strictly speaking, boundedly rational. Their expectation formation mechanism is represented by a VAR model. These expectations would be fully rational if the VAR is the true data generating process. Although I show that the VAR provides an appropriate characterization of the data, it remains only an

---

[2]Source: http://www.econ.yale.edu/~shiller

approximation, which does not take all aspects of the data generating process, such as the existence and strategies of other agents, explicitly into account.

The third type of agent also follows a short-term strategy. Regarding expected price changes, however, this type takes the exact opposite, or contrarian, stance from the rational speculators. These agents are therefore referred to as contrarian speculators, or contrarians. When the rational speculators expect an $x\%$ increase in the price, the contrarians expect an $x\%$ decrease and vice versa.

Several studies provide empirical evidence showing that agents do indeed sometimes embark on such contrarian strategies, (e.g. Kaniel et al., 2008, or Grinblatt and Keloharju, 2000), which is further supported by experimental evidence by Bloomfield et al. (2009). In addition, Park and Sabourian (2011) provide a theoretical justification of contrarian behavior, while Lakonishok et al. (1994), Jegadeesh and Titman (1995), and Dechow and Sloan (1997) discuss the profitability of such strategies. This paper does not provide a theory or intuition for contrarian behavior. Instead, I motivate the existence of contrarians empirically, by showing that observed market dynamics can be replicated rather well when a certain fraction of market participants is forming contrarian expectations. While the existence of rational speculators can explain much of the volatility observed on financial markets, the contrarians turn out to be an essential element of the model in order to approximate observed prices also in terms of correlation.

Contrarian beliefs are in particular helpful in explaining the high valuations that the stock market reached at the end of the 1990s, mainly driven by technology stocks. Whether this episode constituted a bubble has been the subject of debate among many authors, including Ofek and Richardson (2003), Pástor and Veronesi (2006), Bradley et al. (2008), O'Hara (2008) and Phillips et al. (2011). The results in this paper indicate that dividend expectations are not the dominant factor in the observed price increases during the 1990s. In this sense, it could be justified to classify this event as a bubble. Nevertheless, it was not a rational bubble as defined by Blanchard and Watson (1982), since the results show that rational speculators would have driven the market in the opposite direction. Instead, the observed dynamics of the 1990s can

be closely approximated by the contrarian valuation model, suggesting that nonrational beliefs inflated this bubble.

To capture the observed regime switching behavior of financial markets (documented by e.g. Ang and Bekaert, 2002, or Guidolin and Timmermann, 2008), I allow the agents to switch between strategies. Agents are assumed to observe the recent performance of each strategy and choose their own strategy accordingly, following the evolutionary selection scheme introduced by Brock and Hommes (1997, 1998). This scheme has been applied in many theoretical and empirical studies of heterogeneous agent models in finance, including Boswijk et al. (2007), Branch and Evans (2010) and Lof (2013). Similar concepts, in which agents apply learning principles to update expectations are considered by Timmerman (1994), Hong et al. (2007), and Branch and Evans (2011), among others. Hommes et al. (2005) and Bloomfield and Hales (2002) provide experimental evidence in favor of such principles being applied in the formation of expectations. Alternatively, the fractions of different types of agents may be held constant (Szafarz, 2012), or vary according to an exogenous process, such as the business cycle (Lof, 2012a).

As opposed to Brock and Hommes (1997, 1998), the expectations of different agents are in this paper empirically generated by a VAR process. This VAR approach is also recently applied by Cornea et al. (2012) to a heterogeneous agent model of the New Keynesian Philips curve, in which price-setters are allowed to switch between forward-looking and naive backward-looking inflation expectations. Cornea et al. (2012) generate only the expectations of the forward-looking price-setters by a VAR. In this paper, on the other hand, I let all three types of agents form expectations based on the same VAR framework, such that all agents have the same information set. Nevertheless, despite having identical information sets, the agents do not form identical valuations of the asset. Since the expectations are derived from an unrestricted VAR, the valuation based on expected long-term dividends and the valuation based on expected short-term price changes, do not necessarily align.

This paper proceeds as follows. The next section outlines the present value model, the concept of rational bubbles and the log-linear approximation by Campbell and Shiller (1988).

In Section 3.3, the VAR approach is reviewed and applied to three representative agent models, in which the representative agent is either a rational long-term investor, a rational speculator or a contrarian. In Section 3.4, these models are merged into one regime switching model. The section further includes estimation results and specification tests. In Section 3.5, the model is generalized to allow for time-varying discount factors. Section 3.6 concludes.

## 3.2   The present value model and rational bubbles

According to the standard present value model, the price of an asset should equal the discounted present value of the cash flows (dividends) that an asset is expected to generate:

$$P_t = \sum_{i=1}^{\infty} \delta^i E_t [D_{t+i}], \qquad (3.1)$$

in which which $P_t$ refers to the asset price and $D_t$ to its underlying dividend. The discount factor $\delta$ is for simplicity assumed to be constant, implying risk-neutrality. In Section 3.5, I examine the validity of this assumption by considering several time-varying discount factors. Assuming rationality and market efficiency requires that the conditional expectation operator $E_t[\cdot]$ is the optimal prediction conditional on all available information. Because in equation (3.1), the value is entirely determined by expected dividends, or fundamentals, this expression is sometimes referred to as the fundamental value which would be equal to the observed market price if all agents are rational fundamentalists (e.g. Szafarz, 2012).

Agents are not necessarily planning to hold the asset for a long period and may be more interested in the short-term trading profits rather than long-term dividend yields. If agents are planning to hold the asset for a short time only, say one period, the value of the asset should equal the discounted sum of the expected dividend paid out in the next period and the expected price at which the asset can be sold subsequently:

$$P_t = \delta E_t [P_{t+1} + D_{t+1}]. \qquad (3.2)$$

47

The long-term model (3.1) is the solution to the short-term model (3.2) under the following transversality condition:

$$\lim_{i \to \infty} \delta^i E_t \left[ P_{t+i} \right] = 0. \tag{3.3}$$

Hence, under this transversality condition the investment horizon of the agents should not have an impact on the price. However, equation (3.2) has a more general solution which does allow for a discrepancy between equations (3.1) and (3.2):

$$P_t = \sum_{i=1}^{\infty} \delta^i E_t \left[ D_{t+i} \right] + C_t, \tag{3.4}$$

in which $C_t \equiv \delta^{-1} C_{t-1}$, or equivalently, $C_t \equiv \delta^{-t} M_t$, in which $M_t$ may be any martingale process (i.e. $E_t \left[ M_{t+1} \right] = M_t$). Because $C_t$ constitutes a discrepancy between the fundamental value and the observed price, it may be referred to as a bubble. However, since the bubble exists due to a violation of the transversality condition rather than the a violation of rationality, Blanchard and Watson (1982) name it a rational bubble. The finding that rational dividend expectations are not sufficiently volatile to explain observed price volatility can be regarded as a rejection of the present value model (3.1) and is in the literature often interpreted as evidence in favor of rational bubbles (Gürkaynak, 2008).

Two recent studies present theoretical analyses of asset pricing models in which long-term fundamentalists and short-term speculators co-exist. Szafarz (2012) finds that the existence of multiple investment horizons is a potential source of price volatility. Anufriev and Bottazzi (2012), however, argue that variation in the investment horizon has a significant effect on market dynamics only when agents hold heterogeneous expectations about future prices. In this paper, I follow an empirical approach by applying the VAR-based tests of present value models by Campbell and Shiller (1987, 1988) to an asset pricing model with heterogeneity in both investment horizons and expectations. As will become evident in the next section, heterogeneity in investment horizons can explain the high level of volatility observed in stock prices. Nevertheless, heterogeneity in expectations appears to be a crucial element required for generating prices that do not only capture the volatility but also obtain a relatively high correlation with

observed stock prices.

Before proceeding to estimation of the VAR it is preferable to apply the log-linear approximation of the present value model derived by Campbell and Shiller (1988). The return on holding an asset for one period ($R_{t+1} = (P_{t+1} + D_{t+1})/P_t$) can be approximated by a linear equation:

$$r_{t+1} = \rho p_{t+1} - p_t + (1 - \rho) d_{t+1} + k, \tag{3.5}$$

in which $p_t \equiv \log(P_t)$, $d_t \equiv \log(D_t)$ and $r_t \equiv \log(R_t)$. The parameter $\rho$ is below, but close to, one: It denotes the mean of the ratio $\frac{P_t}{P_t + D_t}$, which Campbell and Shiller (1988) assume to be approximately constant over time. Following Campbell and Shiller (1988), the constant term $k$ is ignored in much of the analysis below, because explaining price movements rather than levels, is the main objective of this study. Engsted et al. (2012) show by simulation that these log-linear returns are a close approximation to true returns even in the presence of rational bubbles.

The assumption of a constant discount factor as in equations (3.1)-(3.2) implies that expected returns are constant:

$$E_t[R_{t+1}] = \frac{E_t[P_{t+1} + D_{t+1}]}{P_t} = \delta^{-1}. \tag{3.6}$$

Taking conditional expectations on both sides of equation (3.5), substituting constant expected returns ($E_t[r_{t+1}] \equiv \bar{r}$) and re-arranging gives:

$$y_t = \rho E_t[y_{t+1}] + E_t[\triangle d_{t+1}] + k - \bar{r}, \tag{3.7}$$

in which $y_t \equiv p_t - d_t$ denotes the log price-dividend (PD) ratio. Equation (3.7) can be iterated forward to obtain the long-term interpretation of the present value model, in which the valuation of the asset is determined by expected future dividend growth rates:

$$y_t = \sum_{i=0}^{\infty} \rho^i E_t[\triangle d_{t+1+i}] + \frac{k - \bar{r}}{1 - \rho}. \tag{3.8}$$

49

This solution requires the assumption of a transversality condition:

$$\lim_{i \to \infty} \rho^i E_t\left[y_{t+i}\right] \;=\; 0, \tag{3.9}$$

which, like condition (3.3), excludes the possibility of a rational bubble. Equation (3.8) can be interpreted as the log-linear equivalent of (3.1).

It is also possible to derive a short-term interpretation of the log-linear present value model, in which the value of an asset is determined by the expected return of holding the asset for one period. Subtracting $\rho y_t$ from equation (3.7) and dividing both sides by $1 - \rho$ gives:

$$y_t \;=\; \frac{\rho}{1-\rho} E_t\left[\triangle y_{t+1}\right] + \frac{1}{1-\rho} E_t\left[\triangle d_{t+1}\right] + \frac{k-\bar{r}}{1-\rho}, \tag{3.10}$$

or, since $\triangle y_t = \triangle p_t - \triangle d_t$:

$$y_t \;=\; \frac{\rho}{1-\rho} E_t\left[\triangle p_{t+1}\right] + E_t\left[\triangle d_{t+1}\right] + \frac{k-\bar{r}}{1-\rho}. \tag{3.11}$$

In this model the PD ratio is entirely determined by one-period expectations of the change in the price and dividend. Since the parameter $\rho$ is below but close to one, the ratio $\frac{\rho}{1-\rho}$ is a rather large number, implying that the expected price change is the dominant factor in the valuation of the asset. Expectations on future dividends therefore only play a minor role in this short-term valuation model, akin to the models by Hong et al. (2007) and Branch and Evans (2010), in which agents have the option to omit dividends partly or entirely from their expectation formation mechanism. Nevertheless, in this model dividends are not irrelevant, since observed dividends play a role in the VAR-based expectations of price changes.

Unlike the long-term model (3.8), the short-term model (3.11) does not require the transversality condition (3.9) and therefore allows for the existence of a rational bubble. In the next section, I evaluate both models (3.8) and (3.11) using the VAR approach by Campbell and Shiller (1987, 1988).

## 3.3 The VAR approach

Campbell and Shiller (1988) propose to test the log-linear present value model (3.8) based on an estimated VAR($q$) for the log-PD ratio and the dividend growth rate (both measured in logs):

$$v_t \equiv \begin{bmatrix} y_t \\ \triangle d_t \end{bmatrix} = \sum_{i=0}^{q} A_i v_{t-1} + u_t. \tag{3.12}$$

Both the PD ratio and the dividend growth rate are demeaned so that intercept terms are not required and the parameters $k$ and $\bar{r}$ in (3.8) can be disregarded. I estimate a VAR(2) for annual observations of the PD ratio and the dividend growth rate over the period 1872-2011. The lag length of $q = 2$, is selected using the Akaike Information Criterion (AIC). This lag order is consistent with the results of Campbell and Shiller (1988). Table 3.1 depicts the AIC for different lag lengths, as well as diagnostic tests for the selected VAR(2). The second-order VAR seems to describe the data well as there is no sign of autocorrelation or heteroscedasticity in the residuals. Moreover, the results of a Chow forecast test at several potential break points indicate that parameter constancy can not be rejected.

**TABLE 3.1:** VAR specification and diagnostics

| lags | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| AIC | -7.980 | -7.986 | -7.967 | -7.953 | -7.889 | -7.889 |
| Autocorrelation | 17.63 | (0.612) | | | | |
| Heteroscedasticity | 51.62 | (0.231) | | | | |
| Breakpoint | 1890 | 1910 | 1930 | 1950 | 1970 | 1990 |
| Chow FC | 0.578 | 0.403 | 0.345 | 0.998 | 0.976 | 0.624 |

*Notes:* VAR(q) model (3.12), with annual data for 1872-2011. Top: Lag selection based on Akaike information criterion. Middle: LM-type test statistics (p-values in parentheses) for Autocorrelation (5 lags) and Multivariate ARCH (5 lags) in residuals of VAR(2). Bottom: P-values for Chow forecast test for parameter constancy. All three diagnostic tests are executed with JMulti (Lütkepohl and Krätzig, 2004)

In order to proceed with testing the present value model, it is convenient to consider the VAR(2) model in its companion form:

$$\begin{bmatrix} v_t \\ v_{t-1} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ I_2 & O_{2,2} \end{bmatrix} \begin{bmatrix} v_{t-1} \\ v_{t-2} \end{bmatrix} + \begin{bmatrix} u_t \\ O_{2,1} \end{bmatrix}, \tag{3.13}$$

51

or:

$$z_t = Bz_{t-1} + \varepsilon_t, \tag{3.14}$$

in which $z_t \equiv (v_t, v_{t-1})'$. If this VAR provides an accurate description of the data, which the diagnostics in Table 3.1 indeed suggest, the matrix of estimated parameters $B$ can be used to replicate the conditional expectations in equation (3.8), and to compute a time-series of theoretical PD ratios:

$$
\begin{aligned}
y_t^{rl} &= \sum_{i=0}^{\infty} \rho^i E_t \left[ \triangle d_{t+1+i} \right] = \sum_{i=0}^{\infty} \rho^i \left( e_2' B^i z_t \right) \\
&= e_2' B \left( I - \rho B \right)^{-1} z_t,
\end{aligned}
\tag{3.15}
$$

in which $e_i$ is a vector of zeros in which the $i$th element is replaced by one. A full derivation is provided Campbell and Shiller (1988). The superscript $rl$ to the theoretical PD ratio indicates rational and long-term. The generated theoretical PD ratio can be interpreted as an estimate of how the PD ratio would behave if all agents are rational long-term investors, that value assets according to rational expectations of future dividends.

For now, the parameter $\rho$ is calibrated at a fixed value, as in Campbell and Shiller (1988). I set $\rho = 0.958$ which is the sample average of the ratio $\frac{P_t}{P_t + D_t}$. At the end of this section, I discuss the sensitivity of the results with respect to this calibration.

Figure 3.1 shows the theoretical PD ratio ($y_t^{rl}$), as well as the realized PD ratio ($y_t$). The figure looks similar to the charts in Campbell and Shiller (1987). The theoretical PD ratio is quite strongly correlated with the realized PD ratio ($corr \left( y_t^{rl}, y_t \right) = 0.799$), but the volatility of the theoretical PD ratio falls far behind of observed volatility. This is illustrated by the volatility ratio ($\sigma \left( y_t^{rl} \right) / \sigma \left( y_t \right) = 0.135$), which expresses the standard deviation of the theoretical PD ratio as a fraction of the standard deviation of the realized PD ratio. The long-term present value model (3.15) therefore seems able to explain the direction of the stock market, but lacks explanatory power regarding the observed volatility of the stock market. Already in the 1980s, Campbell and Shiller, among others, interpreted this excess volatility as a rejection of present value models. In fact, as Figure 3.1 shows, the discrepancy between the theoretical and observed PD ratio has only increased further since then, with an unprecedented rise in the PD ratio during
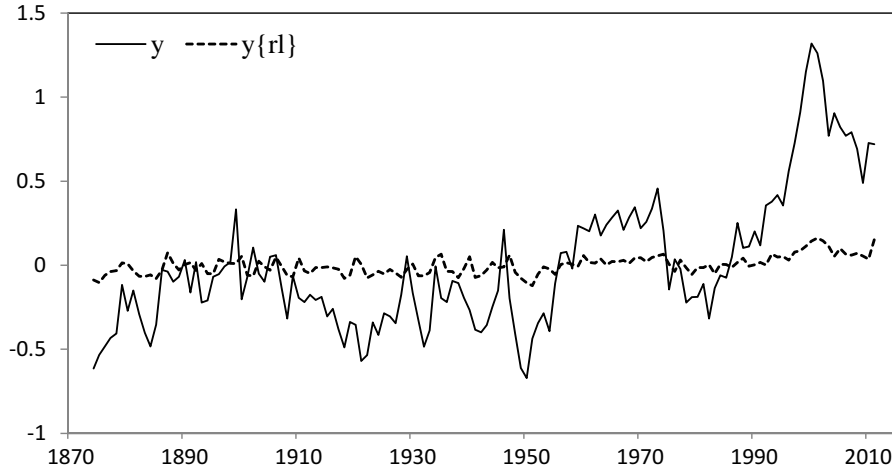
Figure 3.1: Observed PD ratio ($y_t$) and theoretical PD ratio ($y_t^{rl}$), from long-term model (3.15), with $\rho = 0.958$. $corr\left(y_t, y_t^{rl}\right) = 0.799$. $\dfrac{\sigma\left(y_t^{rl}\right)}{\sigma\left(y_t\right)} = 0.135$.

the 1990s, which the present value model fails to capture.

The VAR approach can also be applied to the short-term model (3.11), which is the correct model if all agents are rational speculators. These agents are speculators, as they are mainly interested in short-term trading profits rather than in the dividends the asset generates in the long run. They can be considered (boundedly) rational, however, as they form expectations using the same information set and VAR model as the long-term investors considered above. The conditional expectations of these rational speculators (*rs*) can therefore be replicated based on the estimated VAR, similar as above:

$$y_t^{rs} = \frac{\rho}{1-\rho} E_t\left[\triangle p_{t+1}\right] + E_t\left[\triangle d_{t+1}\right], \tag{3.16}$$

in which:

$$E_t\left[\triangle d_{t+1}\right] = e_2' B z_t, \tag{3.17}$$

and:

$$\begin{aligned} E_t\left[\triangle p_{t+1}\right] &= E_t\left[\triangle y_{t+1}\right] + E_t\left[\triangle d_{t+1}\right] \\ &= E_t\left[y_{t+1}\right] - y_t + E_t\left[\triangle d_{t+1}\right] \\ &= e_1'\left(B-I\right)z_t + e_2' B z_t. \end{aligned} \tag{3.18}$$

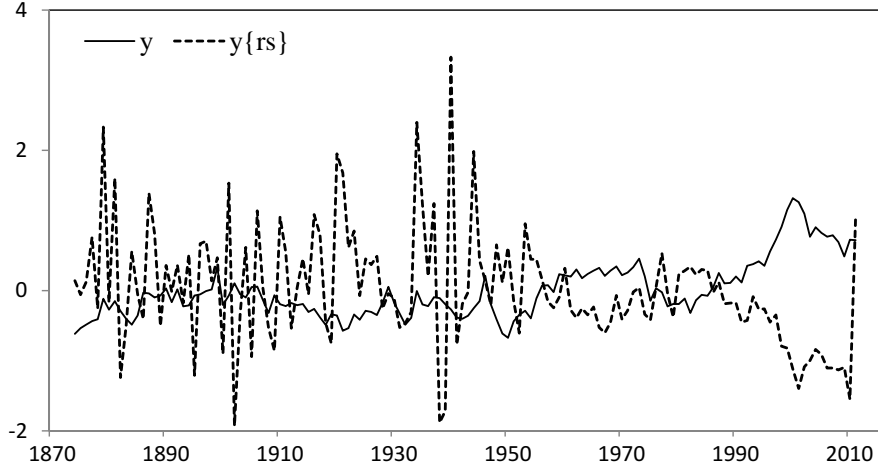In addition, I consider the valuation model according to a second type of speculator: Contrarian

53

Figure 3.2: Observed PD ratio ($y_t$) and theoretical PD ratio ($y_t^{rs}$), from rational speculative model (3.16), with $\rho = 0.958$. $corr(y_t, y_t^{rs}) = -0.403$. $\dfrac{\sigma(y_t^{rs})}{\sigma(y_t)} = 2.065$.

speculators (*cs*) or simply: Contrarians. These agents agree with the rational agents on expected dividends, but form alternative expectations on expected changes in prices:

$$y_t^{cs} = \frac{\rho}{1-\rho} \tilde{E}_t^{cs}[\triangle p_{t+1}] + E_t[\triangle d_{t+1}]. \tag{3.19}$$

In fact, regarding the expected price change, contrarians take the exact opposite stance from the rational speculators:

$$\tilde{E}_t^{cs}[\triangle p_{t+1}] = -E_t[\triangle p_{t+1}]. \tag{3.20}$$

Figure 3.2 shows $y_t^{rs}$ and $y_t$. The model with rational speculative expectations (3.16) appears able to generate large price fluctuations, with the volatility of the theoretical PD ratio even overshooting observed volatility ($\sigma(y_t^{rs})/\sigma(y_t) = 2.065$). Nevertheless, the correlation with the observed PD ratios is very weak, even negative ($corr(y_t^{rs}, y_t) = -0.403$). From Figure 3.2, it can be seen that during several episodes, most notably the 1990s, the theoretical PD ratio moves in the opposite direction from the observed PD ratio. The rational speculative model (3.16) therefore fails to explain the 1990s bull market any better than the long-term model (3.15) does.

Figure 3.3 shows the empirical need for a model with contrarian expectations. The theoretical PD ratio $y_t^{cs}$, which is generated by model (3.19), nearly matches $y_t^{rs}$ in terms of
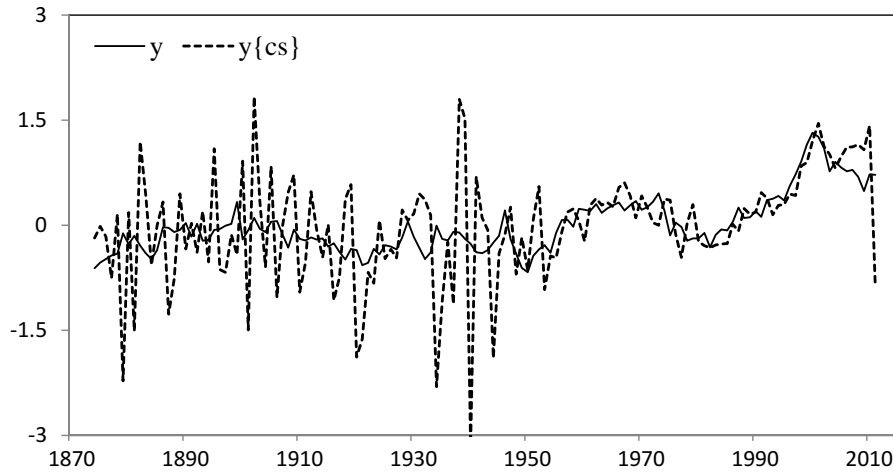
54

Figure 3.3: Observed PD ratio ($y_t$) and theoretical PD ratio ($y_t^{cs}$), from contrarian model (3.19), with $\rho = 0.958$. $corr(y_t, y_t^{cs}) = 0.447$. $\dfrac{\sigma(y_t^{cs})}{\sigma(y_t)} = 1.977$.

volatility: ($\sigma(y_t^{cs})/\sigma(y_t) = 1.977$). Unlike the rational speculative model, however, the contrarian model generates a PD ratio that is positively correlated with the observed PD ratio ($corr(y_t^{cs}, y_t) = 0.447$). Although this correlation remains quite low compared to the long-term model (3.15), it is evident from Figure 3.3 that in recent decades the contrarian model traces the observed PD ratio remarkably well.

Based on Figure 3.1, it can be argued that the bull market in the 1990s was a bubble. It was, however, not a rational bubble, as in that case the rational speculative model (Figure 3.2) should be able to replicate the bubble. Instead, I find that the model requires nonrational, or contrarian, beliefs in order to explain the 1990s bubble.

It is evident from Figures 3.1-3.3 that the performance (or fit) of the three alternative models changes over time, which could indicate misspecification of the VAR, due to the existence of structural breaks or time-varying parameters. The diagnostic tests presented in Table 3.1, however, indicate that the VAR is correctly specified. In addition, I estimate the VAR and generate $y_t^{rl}$, $y_t^{rs}$ and $y_t^{cs}$ again for the last 40 years in the sample only, which are presented in Figure 3.4. These plots tell a roughly similar story as Figures 3.1-3.3, suggesting that the time-varying performance of the three models is not the result of misspecification of the VAR.

Instead, the time-varying fit of the three models could indicate that the market is subject to regime switching behavior, with agents switching between the long-term strategy based on
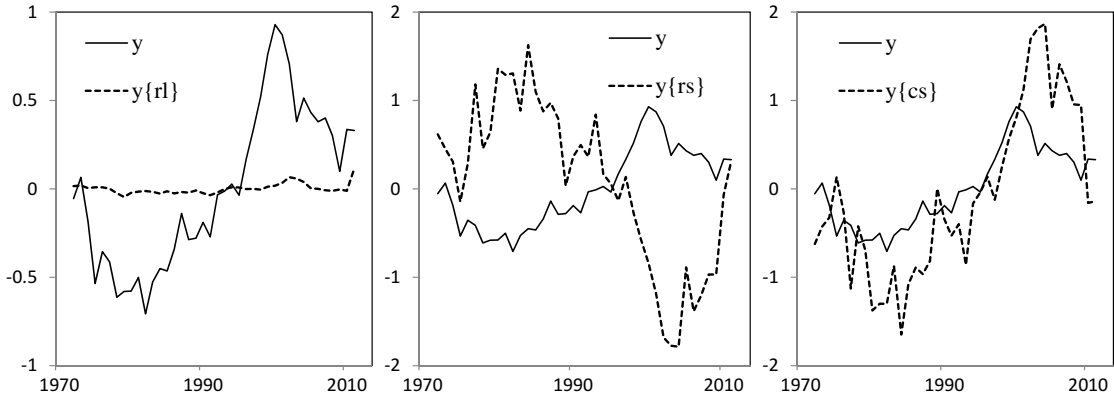
55

Figure 3.4: Observed PD ratio ($y_t$) and theoretical PD ratio ($y_t^{rl}$, $y_t^{rs}$, and $y_t^{cs}$), from models (3.15), (3.16), and (3.19), for 1972-2011.

expected dividends, and more speculative (rational or contrarian) strategies. In the next section, I therefore combine equations (3.15), (3.16) and (3.19) into one regime switching model, in which the asset price is determined by the interaction of rational long-term investors, rational speculators and contrarians.

So far, the parameter $\rho$ is calibrated at the sample average of the ratio $\frac{P_t}{P_t+D_t}$. The obtained results are somewhat sensitive to this calibration. This is illustrated in Figure 3.5, which shows volatility ratios and the correlation between realized and theoretical PD ratios, for different values of $\rho$, for all three models. For the long-run model, the sensitivity with respect to $\rho$ is rather modest. Campbell and Shiller (1988) make the same observation. For the speculative models, however, small changes in $\rho$ do have a great impact. Calibrating $\rho$ and disregarding its uncertainty seems therefore inappropriate. Instead, I estimate $\rho$ in the remainder of this paper
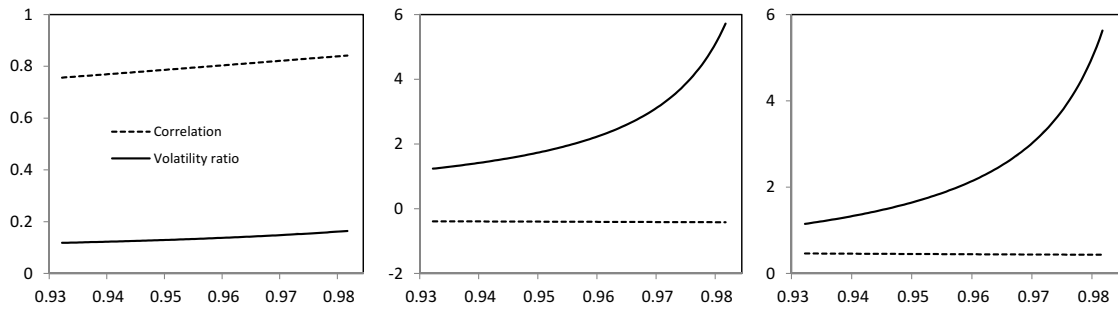


Figure 3.5: $corr\left(y_t, y_t^j\right)$ and $\frac{\sigma\left(y_t^j\right)}{\sigma\left(y_t\right)}$ for different values of $\rho$, for $j = rl$ (left), $j = rs$ (middle) and $j = cs$ (right)

56

jointly with the other parameters in the model.

## 3.4    Heterogeneous agents

The results in the previous section indicate that the long-run present value model (3.15) can explain the direction of stock market movements, but not its excess volatility. The speculative models (3.16) and (3.19) are able to generate sufficient volatility, but their correlation with the observed market falls short of the long-run model. In an attempt to specify a model which is able to capture both correlation and volatility, I consider an economy in which all three agents (long-term rational investors, rational speculators and contrarians) are present:

$$y_t^{ha} = G_t^{rl} y_t^{rl} + G_t^{rs} y_t^{rs} + G_t^{cs} y_t^{cs}, \tag{3.21}$$

in which the subscript *ha* denotes heterogeneous agents. The fractions of each type of agent are denoted by $G_t^{lr}$, $G_t^{sr}$ and $G_t^{sc}$ and are allowed to vary over time. This process of switching between agent types or regimes is modeled based on evolutionary selection following Brock and Hommes (1998), such that the fraction of each type of agents increases when its predictions outperform the other types. The predictions of each type are evaluated by a measure of fitness representing the distance between the theoretical PD ratio and the realized PD ratio in the previous period:

$$U_t^j = -\left(y_{t-1}^j - y_{t-1}\right)^2 \quad j \in \{rl, rs, cs\}. \tag{3.22}$$

The fractions of each type are then determined by multinomial logit probabilities:

$$G_t^j = \frac{\exp\left(\beta^j U_t^j\right)}{\sum_k \exp\left(\beta^k U_t^k\right)} \quad j,k \in \{rl, rs, cs\}, \tag{3.23}$$

such that the fractions of the three types sum to one. The parameters $\beta$ denote the intensity of choice, which indicate the willingness of agents to switch between strategies. While Brock and Hommes (1998) hold $\beta$ constant across types, I allow for type-specific intensities of choice.
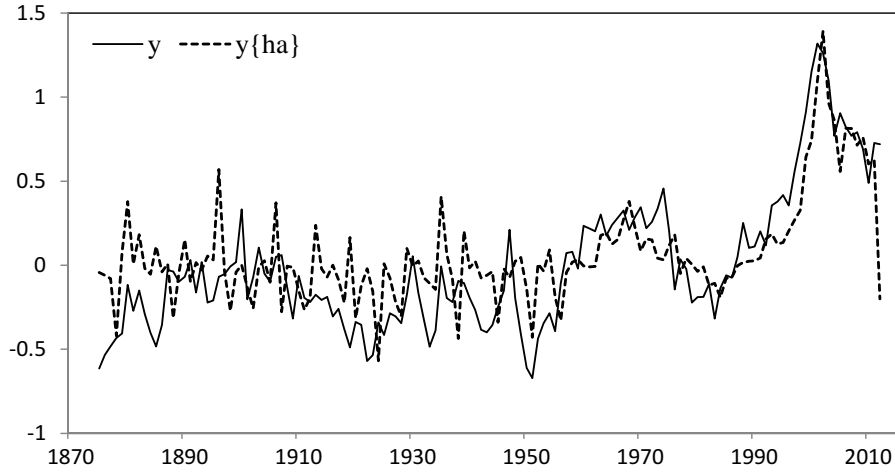
Figure 3.6: Observed PD ratio ($y_t$) and theoretical PD ratio ($y_t^{ha}$), from heterogeneous agent model (3.21), with $\rho$ and $\beta$ estimated by NLS (See Table 3.2). $corr\left(y_t, y_t^{ha}\right) = 0.759$. $\dfrac{\sigma\left(y_t^{ha}\right)}{\sigma\left(y_t\right)} = 0.752$.

This setting accommodates the idea by Hong et al. (2007) that agents may hold heterogeneous thresholds for switching between strategies.

To obtain estimates of $\beta$ and $\rho$, I estimate the following model by nonlinear least squares (NLS):

$$y_t = y_t^{ha} + \varepsilon_t. \tag{3.24}$$

The top row of Table 3.2 shows the parameter estimates, while Figure 3.6 shows a plot of the theoretical PD ratio $y_t^{ha}$. The generated PD ratio is highly correlated with the realized PD ratio; $corr\left(y_t^{ha}, y_t\right) = 0.759$, which is of the same magnitude as the correlation coefficient for the long-term model considered in Section 3.3. The volatility ratio for the heterogeneous agent model is, however, much larger ($\sigma\left(y_t^{ha}\right)/\sigma\left(y_t\right) = 0.752$). Unlike the representative agent models considered in Section 3.3, the heterogeneous agent model is able to explain both the direction as well as the volatility of the observed PD ratio to a large extent.

The fitted values of model (3.24), $\widehat{y_t^{ha}}$, are used to estimate the following regression by OLS:

$$y_t = \phi\widehat{y_t^{ha}} + \varepsilon_t. \tag{3.25}$$

Table 3.2 reports the estimate and standard error of $\phi$, showing that the null hypothesis that $\phi = 1$ can not be rejected.

TABLE 3.2: Estimation results

| | $\rho$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\phi$ | $\sigma(y_t^j)/\sigma(y_t)$ | $corr\left(y_t, y_t^j\right)$ | $R^2$ |
|---|---|---|---|---|---|---|---|---|
| *ha* | 0.966 | 0.799 | 5.175 | 1.125 | 0.962 | 0.752 | 0.759 | 0.548 |
| | (0.004) | (0.599) | (6.156) | (0.401) | (0.029) | | | |
| *rl* | 1.000 | . | . | . | 4.474 | 0.193 | 0.865 | 0.297 |
| | (0.073) | | | | (0.548) | | | |
| *rs* | 0.000 | . | . | . | 3.933 | 0.080 | 0.317 | 0.044 |
| | (0.000) | | | | (0.497) | | | |
| *cs* | 0.000 | . | . | . | 3.933 | 0.080 | 0.317 | 0.044 |
| | (0.202) | | | | (0.568) | | | |

*Notes:* NLS estimates and measures of fit for model (3.21)-(3.24). *ha*: Heterogeneous agents and evolutionary dynamics (3.22)-(3.23). *rl*: $G_t^{rl} = 1$, $G_t^{rs} = G_t^{cs} = 0$. *rs*: $G_t^{rs} = 1$, $G_t^{rl} = G_t^{cs} = 0$. *cs*: $G_t^{cs} = 1$, $G_t^{rl} = G_t^{rs} = 0$. $\phi$ is estimated by model (3.25). Annual data for 1872-2011. Standard errors (in in parentheses) are computed using 10.000 bootstrap replications.

In order to take into account the uncertainty underlying the estimated parameters in the VAR model (3.12), all standard errors in Table 3.2 are based on the following bootstrap procedure:

1. Generate simultaneously an artificial series ($T + 100$ observations) of dividend growth rates from the VAR model (3.12) using the parameter estimates $\widehat{B}$, and an artificial series ($T + 100$ observations) of PD ratios from the model (3.21)-(3.24) using the parameter estimates $\widehat{\beta}$ and $\widehat{\rho}$. The innovations to both series are drawn (with re-sampling) from the fitted residuals $e_2'\widehat{u}_t$ and $\widehat{\varepsilon}_t$.

2. Use the last T observations from both artificial series to estimate models (3.24) and (3.25). Store the estimates $\widetilde{\beta}$, $\widetilde{\rho}$ and $\widetilde{\phi}$.

3. Repeat steps 1 and 2 $R$ times. For each parameter, the standard deviation of the $R$ artificial estimates is reported in Table 3.2 as the parameter's standard error.

For this procedure, I set $T = 138$, equal to the sample size in the estimations, while the number of replications $R = 10.000$.

Figure 3.7 shows the estimated fractions of each type of agent over time. Rational long-term investors are always present in the economy, with their fraction of the total population fluctuating for most of the time between roughly 40% and 100%. After 1950, their fraction stays close to the lower bound of this interval, suggesting that expected dividends have lost relevance as a determinant of asset prices. This is consistent with the finding of decreasing
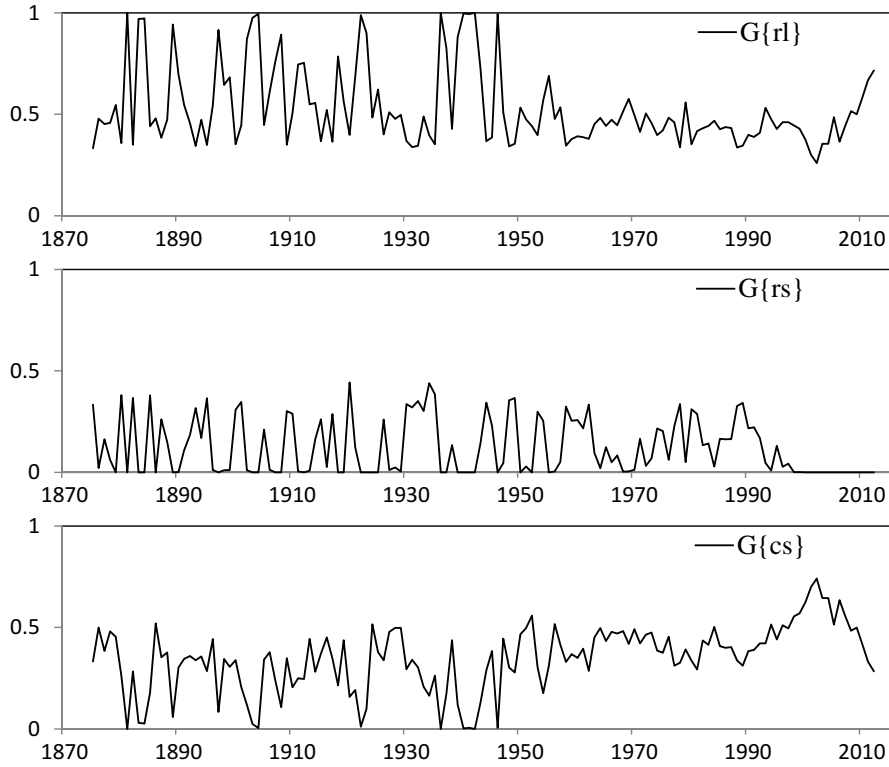
Figure 3.7: Time-varying fractions of long-term investors (top), rational short-term investors (middle) and contrarians (bottom)

dividend yields reported by Fama and French (2001). The fraction of contrarians is relatively high during this period and increases further during the buildup of the 1990s bubble. The fraction of rational speculators stays rather low during the entire sample period.

Table 3.2 further shows estimates of the representative agent models considered in Section 3.3, with the difference that the parameter $\rho$ is now estimated using NLS. These models can be seen as a restricted version of the model (3.21)-(3.24). Instead of the evolutionary dynamics (3.22)-(3.23), the fractions $G_t^{rl}$, $G_t^{rs}$ and $G_t^{cs}$ are restricted to either zero or one. The parameters $\beta$ therefore drop from the model. The correlation coefficients, volatility ratios and $R^2$ reported in Table 3.2 suggest that the heterogeneous agent model is the preferred specification. The long-term model generates a higher correlation coefficient ($corr\left(y_t^{rl}, y_t\right) > corr\left(y_t^{ha}, y_t\right)$) but in all other cases, the heterogeneous agent model generates higher correlation and volatility as well as a better fit in terms of $R^2$. The null hypothesis that $\phi = 1$ is rejected for all three alternatives.

The parameter $\rho$ is estimated under the restriction $0 \leq \rho \leq 1$. For the heterogeneous agent model, the estimate of $\rho$ is rather close to the calibration in Section 3.3. For the representative

agent models, however, a corner solution is reached with $\rho$ estimated at either zero or one. In the log-linear approximation by Campbell and Shiller (1988), the parameter $\rho$ represents the mean of the ratio $\frac{P_t}{P_t+D_t}$. Of course, this mean can never be zero or one as this implies that either prices or dividends are always equal to zero. It is furthermore easy to see that the two speculative models (3.16) and (3.19) reduce to identical models in which one-period dividend expectations are the sole determinant of prices in the case that $\rho = 0$. The finding that highly unrealistic values of $\rho$ are required to obtain the best fit can be interpreted as an economic rejection of the three representative agent models.

For a formal statistical comparison of the heterogeneous agent model and the three representative agent models I rely on the test for nonnested nonlinear regression models developed by Davidson and MacKinnon (1981). The test is based on the following regression:

$$y_t \;\;=\;\; (1-\alpha)y_t^{H1} + \alpha \widehat{y}_t^{H2} + \eta_t, \tag{3.26}$$

in which $y_t^{H1}$ and $y_t^{H2}$ are two nonnested nonlinear regression models, such as the different models considered above. The parameters of $y_t^{H1}$ are estimated jointly with $\alpha$ by NLS, while the test regression further includes the fitted values from NLS estimation of the model $y_t^{H2}$. The hypothesis $H_0: \alpha = 0$ is equivalent to the hypothesis that $y_t^{H1}$ is the correct data generating process. Table 3.3 shows the estimates and standard errors of $\alpha$, from testing $y_t^{ha}$ against $y_t^{rl}$, $y_t^{rs}$ and $y_t^{cs}$ as well as vice-versa. The top row shows the result when $y_t^{H1} = y_t^{ha}$. The hypothesis that $y_t^{ha}$ is correct, can not be rejected against any of the three alternatives. Moreover, the bottom row of Table 3.3 shows that the hypotheses that $y_t^{rl}$, $y_t^{rs}$ or $y_t^{cs}$ are correct are all rejected against the alternative $y_t^{H2} = y_t^{ha}$.

## 3.5   Time-varying discount factors

I have so far assumed a constant discount factor and, as a result, constant expected returns. The log-linear approximation by Campbell and Shiller (1988) does, however, allow for time-varying discount factors. If discount factors are allowed to vary over time, equation (3.7) becomes

**TABLE 3.3:** Nonnested hypothesis tests

| | rl | rs | cs |
|---|---|---|---|
| H1: *ha* | 0.792 | 0.611 | 0.611 |
| | (0.662) | (4.468) | (4.342) |
| H2: *ha* | 0.787 | 0.927 | 0.927 |
| | (0.028) | (0.024) | (0.027) |

*Notes:* NLS estimates of $\alpha$ in model (3.26). Top: $y_t^{H1} = y_t^{ha}$ and $\widehat{y}_t^{H2} = \widehat{y}_t^{j}$, $j \in \{rl, rs, cs\}$ . Bottom: $y_t^{H1} = y_t^{j}$, $j \in \{rl, rs, cs\}$ and $\widehat{y}_t^{H2} = \widehat{y}_t^{ha}$. Rejection of $H_0 : \alpha = 0$ implies rejection of $y_t^{H1}$ (Davidson and MacKinnon, 1981). Annual data for 1872-2011. Standard errors (in parentheses) are computed using 10.000 bootstrap replications.

(disregarding the constant term $k$):

$$y_t \;=\; \rho E_t\left[y_{t+1}\right] + E_t\left[\triangle d_{t+1}\right] - E_t\left[r_{t+1}\right]. \tag{3.27}$$

There are several ways to model time-varying discount factors. Campbell and Shiller (1988) evaluate three simple specifications of the discount factors, based on short-term interest rates, consumption and volatility of the S&P500 index, in addition to a constant discount factor. With a time-varying discount factor, expected returns are computed as follows:

$$E_t\left[r_{t+1}\right] \;=\; \gamma E_t\left[x_{t+1}\right], \tag{3.28}$$

in which $\gamma$ is the risk aversion coefficient and $x_t$ denotes interest rates, consumption or volatility. In the first case, $x_t$ is the log-yield on Treasury Bills (T-bills), representing the opportunity cost of capital. In the second case, $x_t$ is the log-growth rate of consumption, such that the model (3.27) becomes an consumption-based asset pricing model with constant relative-risk aversion utility function. In the third case, $x_t$ is the squared (lagged) log-return of the S&P500 index, as a simple measure of market volatility or risk.

The constant-discount factor is nested in the time-varying specifications. When $\gamma = 0$, it is easily seen that the expected return drops out from equation (3.27), reducing it to the constant discount factor models considered in the previous sections.

I evaluate the three specifications of the time-varying discount factor in the heterogeneous agent model (3.21). Following Campbell and Shiller (1988), I add $x_t$ as a third variable to the

| | $\gamma$ | $\sigma(y_t^j)/\sigma(y_t)$ | $corr\left(y_t, y_t^j\right)$ | $R^2$ |
|---|---|---|---|---|
| constant | . | 0.777 | 0.797 | 0.621 |
| | | | | |
| T-Bill | -0.013 | 0.690 | 0.687 | 0.467 |
| | (0.304) | | | |
| consumption | 0.138 | 0.858 | 0.767 | 0.564 |
| | (0.210) | | | |
| volatility | 0.824 | 0.714 | 0.794 | 0.618 |
| | (0.157) | | | |

*Notes:* NLS estimates and measures of fit for model (3.21)-(3.24), with constant discount factor or time-varying discount factor (3.28) based on interest rates, consumption or volatility. Annual data for 1891-2009. Standard errors (in in parentheses) are computed using 10.000 bootstrap replications.

VAR model (3.12), after which the long-term model (3.15) with time-varying discount factor becomes:

$$y_t^{rl} \quad = \quad \sum_{i=0}^{\infty} \rho^i \left(E_t\left[\triangle d_{t+1+i}\right] - E_t\left[r_{t+i+1}\right]\right) \quad = \quad \left(e_2' - \gamma e_3'\right) B \left(I - \rho B\right)^{-1} z_t, \qquad (3.29)$$

while the speculative models (3.16) and (3.19) become:

$$y_t^{rs} \quad = \quad \frac{\rho}{1-\rho} E_t\left[\triangle p_{t+1}\right] + E_t\left[\triangle d_{t+1}\right] - \frac{1}{1-\rho} E_t\left[r_{t+1}\right], \qquad (3.30)$$

and:

$$y_t^{cs} \quad = \quad \frac{\rho}{1-\rho} \tilde{E}_t^{cs}\left[\triangle p_{t+1}\right] + E_t\left[\triangle d_{t+1}\right] - \frac{1}{1-\rho} E_t\left[r_{t+1}\right], \qquad (3.31)$$

in which:

$$E_t\left[r_{t+1}\right] \quad = \quad \gamma e_3' B z_t. \qquad (3.32)$$

Due to limited data availability, the models with time-varying discount factors can be estimated only for the period 1891-2009. Campbell and Shiller (1988) find that these three time-varying discount factors are not helpful in explaining stock price movements in the long-run model. The results presented in Table 3.4 confirm that this finding also holds for the heterogeneous agent model considered here. Of the four specifications, the constant discount factor is the preferred option. Table 3.4 shows the correlation, volatility ratio and $R^2$ for the estimated heterogeneous

agent models (3.21) with different time-varying discount factors as well as a constant discount factor over this period. The table further shows the NLS estimate of the risk aversion coefficient $\gamma$. Using the discount factor based on either interest rates or consumption, the restriction $\gamma = 0$ (i.e. a constant discount factor) can not be rejected. These specifications are therefore not preferred to the model with constant discount factor. Although the volatility ratio for the consumption-based model is slightly higher than for the model with constant discount factor, the latter yields a higher correlation and a better fit overall.

In the case of a volatility-based discount factor, $\gamma$ is significant, but Table 3.4 shows that also this model is not an improvement in terms of correlation, volatility ratio or $R^2$ with respect to the constant discount factor model. Besides not improving the fit of the model nor increasing the volatility of replicated prices, including a time-varying discount factor based on volatility does not diminish the empirical need for heterogeneous horizons and expectations. As Figure 3.8 shows, with a volatility-based discount factor the estimated fractions of the different types evolve following a similar path as with a constant discount factor (Figure 3.7). In fact, the estimated fraction of contrarians is often even higher than with a constant discount factor.

Various more complex discount factor specifications, besides these three examples, could be considered. As Cochrane (2011) argues, for any behavioral model there exists an equivalent rational expectations model with time-varying discount factor. Nevertheless, this does not imply that modeling discount factors instead of expectations is always the most sensible strategy. The results presented in this paper show that a simple and straightforward extension (allowing for heterogeneous horizons and expectations) can generate significantly more volatility than the linear present value model. Specifying a parametric process for the evolution of a discount factor that is able to accomplishing the same result could instead be a rather complex task. The three specifications considered in this section are at least not adequate.
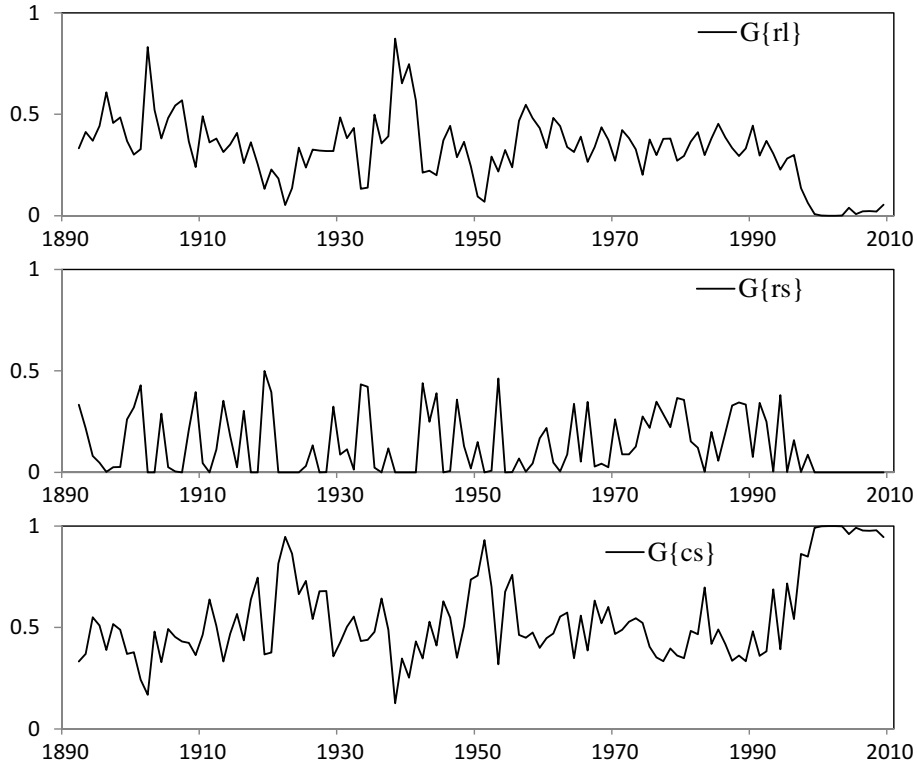
Figure 3.8: Time-varying fractions of long-term investors (top), rational short-term investors (middle) and contrarians (bottom), with volatility-based time-varying discount factor

## 3.6 Conclusion

I develop an empirical asset pricing model in which the expectations of all agents are derived from a VAR representation for price-dividend ratios and dividend growth rates. Taking into account the performance of each strategy in the previous period, agents choose between a long-term strategy, valuing asset based on expected dividends, and two types of short-term strategies, valuing assets mainly based on expected price changes. This heterogeneous agent model is able to generate far more volatile PD ratios than a standard present value model, thereby tackling a considerable part of the excess volatility puzzle.

The existence of speculators can explain the volatility of stock prices. Nevertheless, heterogeneity in expectations among the speculators is required in order to approximate observed prices in terms of volatility as well as correlation. In particular to replicate the stock market during the 1990s accurately, a large fraction of market participants needs to adopt contrarian beliefs. As this requires a deviation from the assumption of rationality, I argue that the 1990s

65

bubble was not a rational bubble.

The introduction of time-varying discount factors into the model does not significantly alter the results. Overall, the results suggest that observed excess volatility with respect to the standard present value model is better explained by nonstandard expectations rather than by time-varying discount factors.

# References

Ang, A. and G. Bekaert: 2002, 'International asset allocation with regime shifts'. *Review of Financial studies* **15**(4), 1137–1187.

Anufriev, M. and G. Bottazzi: 2012, 'Asset Pricing with Heterogeneous Investment Horizons'. *Studies in Nonlinear Dynamics & Econometrics* **16**(4).

Barberis, N., A. Shleifer, and R. Vishny: 1998, 'A model of investor sentiment'. *Journal of financial economics* **49**(3), 307–343.

Barberis, N. and R. Thaler: 2003, 'A survey of behavioral finance'. *Handbook of the Economics of Finance* **1**, 1053–1128.

Blanchard, O. and M. Watson: 1982, 'Bubbles, Rational Expectations and Financial Markets'. *NBER Working Paper* (945).

Bloomfield, R. and J. Hales: 2002, 'Predicting the next step of a random walk: experimental evidence of regime-shifting beliefs'. *Journal of Financial Economics* **65**(3), 397–414.

Bloomfield, R., M. O Hara, and G. Saar: 2009, 'How noise trading affects markets: An experimental analysis'. *Review of Financial Studies* **22**(6), 2275–2302.

Boswijk, H. P., C. H. Hommes, and S. Manzan: 2007, 'Behavioral heterogeneity in stock prices'. *Journal of Economic Dynamics and Control* **31**(6), 1938–1970.

Bradley, D., B. Jordan, and J. Ritter: 2008, 'Analyst behavior following IPOs: the 'bubble period' evidence'. *Review of Financial Studies* **21**(1), 101–133.

Branch, W. and G. Evans: 2010, 'Asset return dynamics and learning'. *Review of Financial Studies* **23**(4), 1651–1680.

Branch, W. and G. Evans: 2011, 'Learning about risk and return: A simple model of bubbles and crashes'. *American Economic Journal: Macroeconomics* **3**(3), 159–191.

Brock, W. A. and C. H. Hommes: 1997, 'A Rational Route to Randomness'. *Econometrica* **65**(5), 1059–1096.

Brock, W. A. and C. H. Hommes: 1998, 'Heterogeneous beliefs and routes to chaos in a simple asset pricing model'. *Journal of Economic Dynamics and Control* **22**(8-9), 1235–1274.

Campbell, J. and R. Shiller: 1987, 'Cointegration and Tests of Present Value Models'. *Journal of Political Economy* **95**(5), 1062–1088.

Campbell, J. Y. and R. J. Shiller: 1988, 'The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors'. *Review of Financial Studies* **1**(3), 195–228.

Cochrane, J. H.: 2011, 'Presidential Address: Discount Rates'. *The Journal of Finance* **66**(4), 1047–1108.

Cornea, A., C. Hommes, and D. Massaro: 2012, 'Behavioral Heterogeneity in US Inflation Dynamics'. *CeNDEF Working paper*.

Davidson, R. and J. MacKinnon: 1981, 'Several tests for model specification in the presence of alternative hypotheses'. *Econometrica* **49**(3), 781–793.

De Long, J., A. Shleifer, L. Summers, and R. Waldmann: 1990a, 'Noise Trader Risk in Financial Markets'. *Journal of Political Economy* **98**(4), 703–738.

De Long, J., A. Shleifer, L. Summers, and R. Waldmann: 1990b, 'Positive Feedback Investment Strategies and Destabilizing Rational Speculation'. *The Journal of Finance* pp. 379–395.

Dechow, P. and R. Sloan: 1997, 'Returns to contrarian investment strategies: Tests of naive expectations hypotheses'. *Journal of Financial Economics* **43**(1), 3–27.

Engsted, T., T. Pedersen, and C. Tanggaard: 2012, 'The Log-Linear Return Approximation, Bubbles, and Predictability'. *Journal of Financial and Quantitative Analysis* **47**(3), 643–665.

Fama, E. F. and K. R. French: 2001, 'Disappearing dividends: changing firm characteristics or lower propensity to pay?'. *Journal of Financial Economics* **60**(1), 3–43.

Gilles, C. and S. LeRoy: 1991, 'Econometric aspects of the variance-bounds tests: A survey'. *Review of Financial Studies* **4**(4), 753–791.

Grinblatt, M. and M. Keloharju: 2000, 'The investment behavior and performance of various investor types: a study of Finland's unique data set'. *Journal of Financial Economics* **55**(1), 43–67.

Guidolin, M. and A. Timmermann: 2008, 'International asset allocation under regime switching, skew, and kurtosis preferences'. *Review of Financial Studies* **21**(2), 889–935.

Gürkaynak, R.: 2008, 'Econometric Tests of Asset Price Bubbles: Taking Stock'. *Journal of Economic Surveys* **22**(1), 166–186.

Hirshleifer, D.: 2001, 'Investor psychology and asset pricing'. *The Journal of Finance* **56**(4), 1533–1597.

Hommes, C., J. Sonnemans, J. Tuinstra, and H. Van de Velden: 2005, 'Coordination of expectations in asset pricing experiments'. *Review of Financial Studies* **18**(3), 955–980.

Hong, H. and J. Stein: 1999, 'A unified theory of underreaction, momentum trading, and overreaction in asset markets'. *The Journal of Finance* **54**(6), 2143–2184.

Hong, H., J. Stein, and J. Yu: 2007, 'Simple forecasts and paradigm shifts'. *The Journal of Finance* **62**(3), 1207–1242.

Jegadeesh, N. and S. Titman: 1995, 'Overreaction, delayed reaction, and contrarian profits'. *Review of Financial Studies* **8**(4), 973–993.

Kaniel, R., G. Saar, and S. Titman: 2008, 'Individual investor trading and stock returns'. *The Journal of Finance* **63**(1), 273–310.

Lakonishok, J., A. Shleifer, and R. Vishny: 1994, 'Contrarian Investment, Extrapolation, and Risk'. *The Journal of Finance* pp. 1541–1578.

LeRoy, S. F. and R. D. Porter: 1981, 'The Present-Value Relation: Tests Based on Implied Variance Bounds'. *Econometrica* **49**(3), 555–74.

Lof, M.: 2012a, 'Heterogeneity in stock prices: A STAR model with multivariate transition function'. *Journal of Economic Dynamics and Control* **36**(12), 1845 – 1854.

Lof, M.: 2012b, 'Rational Speculators, Contrarians and Excess Volatility'. *HECER Discussion Paper* (358).

Lof, M.: 2013, 'Noncausality and Asset Pricing'. *Studies in Nonlinear Dynamics and Econometrics* **17**(2), 211–220.

Lütkepohl, H. and M. Krätzig: 2004, *Applied time series econometrics*. Cambridge University Press.

Ofek, E. and M. Richardson: 2003, 'DotCom Mania: The Rise and Fall of Internet Stock Prices'. *The Journal of Finance* **58**(3), 1113–1138.

O'Hara, M.: 2008, 'Bubbles: Some perspectives (and loose talk) from history'. *Review of Financial Studies* **21**(1), 11–17.

Park, A. and H. Sabourian: 2011, 'Herding and contrarian behavior in financial markets'. *Econometrica* **79**(4), 973–1026.

Pástor, L. and P. Veronesi: 2006, 'Was there a Nasdaq bubble in the late 1990s?'. *Journal of Financial Economics* **81**(1), 61–100.

Phillips, P., Y. Wu, and J. Yu: 2011, 'Explosive Behavior in the 1990s' NASDAQ: When Did Exuberance Escalate Asset Values?'. *International economic review* **52**(1), 201–226.

Shiller, R.: 2003, 'From efficient markets theory to behavioral finance'. *The Journal of Economic Perspectives* **17**(1), 83–104.

Shiller, R. J.: 1981, 'Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?'. *American Economic Review* **71**(3), 421–36.

Szafarz, A.: 2012, 'Financial crises in efficient markets: How fundamentalists fuel volatility'. *Journal of Banking and Finance* **36**(1), 105 – 111.

Timmerman, A.: 1994, 'Can agents learn to rational expectations? Some results on convergence and stability of learning in the UK stock market'. *The Economic Journal* pp. 777–797.

West, K. D.: 1988, 'Dividend Innovations and Stock Price Volatility'. *Econometrica* **56**(1), 37–61.

# Chapter 4

# Noncausality and asset pricing[1]

## 4.1  Introduction

Recent research (e.g. Lanne and Saikkonen, 2011a,b) finds that many financial and economic variables are noncausal, in the sense that when these variables are modeled as linear autoregressions, current observations seem to depend on both past and future realizations, rather than only on past realizations. This paper discusses noncausality of asset prices and dividends. Recent literature dealing with noncausality focuses mainly on econometric issues, such as instrument selection in GMM estimation (Lanne and Saikkonen, 2011a) and forecasting (Lanne et al. 2012 a,b). In this paper the focus is not on empirical implications but rather on the economic interpretation of noncausality. I show by simulation that noncausality is observed when relevant information is excluded from the econometric model. Asset prices are shown to be noncausal when the econometric model is based on observed market data, but fails to include the correct expectation formation mechanism.

A noncausal autoregressive (AR) process differs from a conventional causal AR process in the dependence on both future and past errors, implying that future errors are predictable given the realized observations of the variable in question. An early discussion of noncausal autoregressions is provided by Breidt et al. (1991). Recently, Lanne and Saikkonen (2011b)

---

[1]This chapter is based on an article published in *Studies in Nonlinear Dynamics and Econometrics* (Lof, 2013)

introduced a useful reparametrization of the noncausal AR process allowing for explicit dependence on both leads and lags of the variable in question. A stationary noncausal AR($r$,$s$) process $y_t$, depending on $r$ lags and $s$ leads (with $r$ and $s$ both positive integers), is defined by:

$$\phi(L)\varphi(L^{-1})y_t = \varepsilon_t, \tag{4.1}$$

with $\phi(L) = 1 - \phi_1 L - ...\phi_r L^r$, $\varphi(L^{-1}) = 1 - \varphi_1 L^{-1} - ...\varphi_r L^{-s}$, $\varepsilon_t \sim i.i.d.(0, \sigma^2)$ and $L$ is a standard lag operator ($L^k y_t = y_{t-k}$). Both polynomials have their roots outside the unit circle. If $\varphi_j \neq 0$, for some $j \in \{1,..,s\}$, (4.1) is a noncausal process, which may be referred to as purely noncausal if $\phi_1 = ... = \phi_p = 0$. When $y_t$ is a vector, (4.1) defines a noncausal vector autoregressive process VAR($r$,$s$) (Lanne and Saikkonen, 2013).

Lanne and Saikkonen (2011b) point out that noncausality is related to noninvertibility, as noncausal AR processes and noninvertible Moving Average (MA) processes are close approximations of each other. Exact definitions of causal and invertible processes are provided by Brockwell and Davis (1991) or Meitz and Saikkonen (2013): An ARMA process is invertible when the error term can be expressed as a weighted sum of past and present components of the process: $\varepsilon_t = \sum_{j=0}^{\infty} \alpha_j y_{t-j}$, with $\sum_{j=0}^{\infty} |\alpha_j| < \infty$. An ARMA process is causal when each component can be expressed as a weighted sum of past and present error terms. For example, it is well known that any stationary causal AR($r$,0) process has a backward-looking, infinite-order, MA representation:

$$y_t = \phi(L)^{-1}\varepsilon_t = \sum_{j=0}^{\infty} \mu_j \varepsilon_{t-j}, \tag{4.2}$$

in which $\sum_{j=0}^{\infty} \mu_j z^j = \mu(z) \equiv \phi(z)^{-1}$. The MA representation of a purely noncausal AR(0,$s$) process is, on the other hand, forward-looking:

$$y_t = \varphi(L^{-1})^{-1}\varepsilon_t = \sum_{j=0}^{\infty} \omega_j \varepsilon_{t+j}, \tag{4.3}$$

in which $\sum_{j=0}^{\infty} \omega_j z^{-j} = \omega(z^{-1}) \equiv \varphi(z^{-1})^{-1}$. A noncausal AR($r,s$) process, with $r$ and $s$ both greater than zero, has a MA representation that is both backward- and forward-looking:

$$y_t = \varphi(L^{-1})^{-1} \phi(L)^{-1} \varepsilon_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \tag{4.4}$$

in which $\psi_j$ is the coefficient of $z^j$ in the Laurent-series expansion of $\varphi(z^{-1})^{-1}\phi(z)^{-1}$ (Lanne and Saikkonen, 2011b). Since a stationary noncausal process can not be inverted into a backward-looking MA representation, its errors are nonfundamental[2]. Nonfundamentalness arises when the agents in the economy base their expectations on a larger information set than the information set available to an econometrician, in which case the residuals from the estimated autoregression are not an interpretable function of the true shocks to the agents' information (Hansen and Sargent, 1991; Alessi et al., 2011). In this situation, a noncausal autoregression may fit the data better, because it takes the omitted information into account, by allowing for predictable errors, even without explicit specification of the correct information set (Lanne and Saikkonen, 2011b)[3].

The agents' information set is a flexible concept. The most obvious example of an econometrician having a smaller information set than the agents in the economy is the omission of one or more relevant decision variables from the estimated model. In this paper, I argue that another example of such a situation occurs when the econometrician and the agents observe the same variables, but the econometrician misunderstands the complexity of the expectation formation mechanism, by estimating a linear model while the true mechanism is nonlinear.

Throughout this paper, an observed variable or vector of variables is referred to as noncausal, when a noncausal linear (vector) autoregressive model fits the data better than a causal (vector) autoregressive model. Observed noncausality may be the result of omitted information rather than an actual dependence on future observations. In Section 4.3, I show that noncausality is often observed when a linear univariate autoregressive model is estimated for a variable that

---

[2]This paper only deals with stationary time-series, excluding the 'borderline' possibility of a unit root process that is not invertible but fundamental (Alessi et al., 2011).

[3]Forni et al. (2009) propose an alternative approach by applying large-dimensional factor models, which increase the econometrician's information set and thereby avoid nonfundamentalness.

was actually generated by a multivariate or nonlinear process. In section 4.4, the existence of heterogeneous beliefs is shown to be a possible source of noncausality of asset prices. In this case, different agents form different expectations about the future, making it difficult for an econometrician to observe or infer these expectations. This is an important missing piece of information, since on financial markets these expectations ultimately drive asset prices.

To motivate the search for sources of noncausality in asset pricing, the next section presents empirical evidence that historical US stock prices are indeed noncausal.

## 4.2 Empirical results

To determine whether a causal or noncausal autoregression fits a certain variable $y_t$ better, I will follow the model selection procedure proposed by Lanne and Saikkonen (2011b). First, a causal autoregression AR($p$) is estimated by least squares to find the optimal number of lags $p$ such that the model seems adequate in describing the autocorrelation. In this paper the number of lags is selected by the Bayesian Information Criterion (BIC). Next, model (4.1) is estimated by maximum likelihood (ML) for all possible combinations of $r$ and $s$ for which $r + s = p$, using the ML estimator proposed by Lanne and Saikkonen (2013, 2011b) for univariate and multivariate processes. After estimating all possible AR($r, s$) models, the specification yielding the largest value of the likelihood function is chosen as the adequate autoregression. If for this model $s > 0$, the variable $y_t$ is referred to as noncausal.

The noncausal process as defined in equation (4.1) does not require any distributional assumptions, except that the errors are *i.i.d.* Estimating the model, however, does require further assumptions on the distribution. Causal and noncausal autoregressive processes are indistinguishable when the error terms are Gaussian (Breidt et al., 1991). Therefore, a non-Gaussian distribution needs to be assumed. With macro-economic and financial time series this does not need to be a problem, since Gaussianity if often rejected for these time series due to fat tails. In their empirical applications, Lanne and Saikkonen (2013, 2011b) assume t-distributed errors. I follow this assumption. In the empirical results below, this assumption is justified by a test

for normality. For the simulation exercises later in the paper, random errors are drawn from a t-distribution.

The model selection procedure of Lanne and Saikkonen (2011b) is applied to univariate and bivariate time series related to asset pricing, using long-term data on the US stock market provided by Shiller (2005). This dataset includes annual observations from 1871 to 2010 on the value of the S&P500 index ($P_t$) and the average dividends ($D_t$) paid to investors holding shares in this index. Noncausality is checked for the log-difference of prices ($\triangle p_t = \log(P_t) - \log(P_{t-1})$) and dividends ($\triangle d_t = \log(D_t) - \log(D_{t-1})$), as well as for the bivariate processes ($\triangle p_t, \triangle d_t)'$ and $(\delta_t, \triangle d_t)'$, with $\delta_t = \log(P_t/D_t)$ is the log price-dividend (PD) ratio. Table 4.1 depicts the log-likelihood values for all estimated AR($r,s$) models. Log-differenced dividends are found to be causal, but log-differenced prices and both VARs are best described by noncausal models.

Table 4.1 further shows some diagnostic test results. After selecting the number of lags $p$ based on a Gaussian causal AR, Gaussianity of the residuals is tested. Gaussianity is rejected by a Jarque-Bera test for all ARs, justifying estimation by non-Gaussian maximum likelihood. The residuals of the autoregression selected as adequate are furthermore subjected to tests for autocorrelation (Ljung-Box) and conditional heteroscedasticity (McLeod-Li). There is no evidence for remaining autocorrelation or heteroscedasticity at the 5% level. In general, the selected noncausal autoregressions seem to describe these time series well.

**Table 4.1**

|  | $\triangle p_t$ | | $\triangle d_t$ | | $(\delta_t, \triangle d_t)'$ | | $(\triangle p_t, \triangle d_t)'$ | |
|---|---|---|---|---|---|---|---|---|
|  | $(r,s)$ | $\mathscr{L}$ | $(r,s)$ | $\mathscr{L}$ | $(r,s)$ | $\mathscr{L}$ | $(r,s)$ | $\mathscr{L}$ |
|  | *(1,0)* | 41.8 | *(1,0)* | **123.3** | *(2,0)* | -240 | *(1,0)* | -360 |
|  | *(0,1)* | **42.8** | *(0,1)* | 119.9 | *(1,1)* | **-228** | *(0,1)* | **-350** |
|  |  |  |  |  | *(0,2)* | -229 |  |  |
| JB | 0.01 | | 0.00 | | 0.00 | | 0.00 | |
| LB | 0.08 | | 0.20 | | 0.19 | 0.22 | 0.13 | 0.27 |
| MLL | 0.36 | | 0.12 | | 0.11 | 0.06 | 0.08 | 0.06 |

*Notes:* Log-likelihood values for all possible AR($r,s$) specifications such that $p = r + s$. The specification that maximizes the log-likelihood for each variable is depicted in bold. The lag length $p$ is selected by the BIC, based on a causal Gaussian AR, after which Gaussianity of the residuals is tested with a Jarque-Bera test. JB refers to the p-value of this test. LB and MLL refer to the p-values of the Ljung-Box and McLeod-Li tests (5 lags), applied to the residuals of the optimal (non)causal t-distributed AR.

The VAR including PD ratios and dividends $(\delta_t, \triangle d_t)'$ was proposed by Campbell and Shiller (1988) to model agents' expectations of PD ratios and dividends under constant discount rates. The result that $(\delta_t, \triangle d_t)'$ is noncausal is consistent with findings by Lanne and Saikkonen (2013), who show that the VAR proposed by Campbell and Shiller (1987) to model the expected term spread of interest rates is also noncausal. Noncausality of $(\delta_t, \triangle d_t)'$ implies that agents do not base their expectations only on lags of the PD ratio and the dividend growth rate. The same argument applies to the second VAR in Table 4.1, including the growth rates of prices and dividends $(\triangle p_t, \triangle d_t)'$. Taking expectations conditional on all information dated $t-1$ and earlier shows that these expectations can not be expressed as a function of observable data alone:

$$
E_{t-1}\begin{bmatrix} \delta_t \\ \triangle d_t \end{bmatrix} = \Phi_1 \begin{bmatrix} \delta_{t-1} \\ \triangle d_{t-1} \end{bmatrix} + \Pi_1 E_{t-1} \begin{bmatrix} \delta_{t+1} \\ \triangle d_{t+1} \end{bmatrix} + E_{t-1} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}
$$

$$
E_{t-1}\begin{bmatrix} \triangle p_t \\ \triangle d_t \end{bmatrix} = \Pi_1 E_{t-1} \begin{bmatrix} \triangle p_{t+1} \\ \triangle d_{t+1} \end{bmatrix} + E_{t-1} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}.
$$

An economic interpretation of noncausality is therefore that agents' expectations are not revealed when only realized prices and dividends are observed. Future realizations or a wider information set are required to infer the true expectations. This observed dependence on leading observations may be caused by misspecification of the agents' information set. This issue is further discussed in the remainder of this paper.

## 4.3 Misspecified autoregressions

By simulating two simple AR processes, I illustrate that misspecification of the econometric model can cause noncausality. In the first example the variable of interest is generated as a multivariate model, but estimated as a univariate process. In the second example the data generating process is nonlinear, while a linear model is estimated.

First, the omitted-variable problem is considered. The data are generated by a first order causal bivariate process:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{x,t} \\ \varepsilon_{y,t} \end{bmatrix} \qquad \varepsilon_{x,t}, \varepsilon_{y,t} \sim t_3(0,1). \tag{4.5}$$

The *i.i.d.* errors $\varepsilon_{x,t}$ and $\varepsilon_{y,t}$ t-distributed with three degrees of freedom, zero mean and variance one. The simulated errors are t-distributed rather than Gaussian, because Gaussian causal and noncausal ARs are indistinguishable, as discussed in Section 4.2. I calibrate $a = c = 0.8$ and generate 200 observations of $x_t$ and $y_t$ for different values of $b$. After this simulation, $y_t$ is dropped from the information set and $x_t$ is estimated as a univariate AR process to check non-causality by the model selection procedure discussed in the previous section. This simulation is repeated 5000 times. Table 4.2 shows how often the model selection procedure selects causal and noncausal representations for different values of $b$.[4] When $b = 0$, the causal autoregression is the correct specification and is selected in 98% of the simulations. However, when $b \neq 0$, $x_t$ is driven by two shocks $\varepsilon_{x,t}$ and $\varepsilon_{y,t}$, while only one shock can be identified by estimating an autoregression. Due to this nonfundamentalness, a noncausal autoregression is selected as the adequate specification more often, up to 40% of the simulations for $b = 0.8$. Interestingly, when $b$ becomes larger in absolute value, $\varepsilon_{y,t}$ becomes the dominant shock and the causal AR is again selected more often. In the case that $b = 10$, the contribution of $\varepsilon_{x,t}$ to the dynamics of $x_t$, relative to the contribution of $\varepsilon_{y,t}$, is so small that the true process can be well approximated by a causal AR process with only one shock.

**TABLE 4.2**

| $b$ | -10 | -0.5 | 0 | 0.2 | 0.5 | 0.8 | 1 | 2 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| Causal | 93% | 68% | 98% | 94% | 69% | 60% | 66% | 87% | 93% |
| Noncausal | 7% | 32% | 2% | 6% | 31% | 40% | 34% | 13% | 7% |

*Notes:* Percentage of causal and noncausal outcomes of the AR for $x_t$ after 5000 simulations of model (4.5), with $a = c = 0.8$ and different values of $b$. The sample size in each simulation is 200 observations.

---

[4]The simulations are also carried out for different values of $a$ and $c$ between -1 and 1 and for different sample sizes (500 and 1000). As long as $a$ and $c$ are not too close to zero, (i.e. the simulated data are not white noise), the results are similar to those in Table 4.2 and are therefore not explicitly reported.

Next, a univariate nonlinear Logistic Smooth Transition Autoregressive (LSTAR) process is generated:

$$y_t = \alpha_1 y_{t-1}(1 - G(s_{t-1})) + \alpha_2 y_{t-1} G(s_{t-1}) + \varepsilon_t$$
$$G(s_{t-1}) = (1 + exp[-\gamma s_{t-1}])^{-1}$$
$$\varepsilon_t \sim t_3(0, 1). \qquad (4.6)$$

This process is a weighted average of two causal AR(1) regimes. Since the weights are time-varying, the process is nonlinear. However, when $\gamma = 0$, the transition function $G(s_{t-1}) = 1/2$ in all periods, so the process is linear. On the other hand, when $\gamma = \infty$, $G(s_{t-1})$ is either zero or one, meaning the process reduces to a Threshold Autoregressive (TAR) process. In short, the process becomes more nonlinear when $\gamma$ increases. I choose the transition variable $s_{t-1} = \triangle y_{t-1}$ and the calibration $\alpha_1 = 0.8$ and $\alpha_2 = -0.2$, so that each regime is stationary and differs considerably from the other regime. A sample of 200 observations is simulated for different values of $\gamma$: 0, 0.2, 0.5, 1, 2 and 10.000($\approx \infty$), after which a linear AR model is fitted to the data to check for noncausality. Table 4.3 displays the results of 5000 repetitions. In the linear case ($\gamma = 0$), a noncausal specification is selected in 4% of the simulations. However, the number of noncausal representations selected steadily increases with $\gamma$, up to 66% of the simulations for the TAR model. These results show that not only after omitting variables, but also after misspecification of the functional form, a noncausal process often approximates the true process better than a causal process, even if the true process depends by no means on the future.

**TABLE 4.3**

| $\gamma$ | 0 | 0.2 | 0.5 | 1 | 2 | $\infty$ |
|---|---|---|---|---|---|---|
| Causal | 96% | 92% | 82% | 68% | 53% | 34% |
| Noncausal | 4% | 8% | 18% | 33% | 47% | 66% |

*Notes:* Percentage of causal and noncausal outcomes of the AR for $y_t$ after 5000 simulations of model (4.6), with $s_{t-1} = \triangle y_{t-1}$, $\alpha_1 = 0.8$, $\alpha_2 = -0.2$. The sample size in each simulation is 200 observations.

## 4.4 Heterogeneous expectations

Returning to asset pricing, the results of the previous section suggest that the observed non-causality in Table 4.1 could be the result of misspecification: The evolution of asset prices over

time depends on information that may be known to the agents, but is not observable by an econometrician.

The existence of heterogeneous beliefs is a natural candidate for such a situation. Kasa et al. (2010) derive conditions under which informational heterogeneity (agents receiving different signals about future dividends) imposes agents to forecast the forecasts of other agents, as in Townsend (1983), which leads to a nonrevealing equilibrium. Kasa et al. (2010) explicitly show how the process of prices and dividends is under these conditions not invertible into a backward-looking moving average process and argue that an econometrician who does not observe these different signals will misinterpret the (nonfundamental) residuals from a VAR as shocks to the agents' information.

To check what type of investor behavior generates noncausality, I simulate asset prices under different expectation regimes. I consider a representative-agent model and two models featuring boundedly rational agents with heterogeneous beliefs. After each simulation, I act as an econometrician who does not understand the structure of the underlying model and estimate both causal and noncausal VARs for prices and dividends, to find out which VAR fits the data best. The starting point for this simulation exercise are the dividends, which are assumed to be exogenous, not depending on asset prices. To be precise, dividends are generated by a causal AR(1) process:

$$d_t = \alpha_1 + \alpha_2 d_{t-1} + \varepsilon_t, \tag{4.7}$$

with $\varepsilon_t \sim t_3(0, \sigma_\varepsilon^2)$. The fundamental value $p_t^*$ of the asset equals the sum of all expected future dividends, discounted at a constant discount factor $r$:

$$
\begin{aligned}
p_t^* &= \sum_{i=1}^{\infty} \frac{E_{t-1}[d_{t+i}]}{(1+r)^i} \\
E_{t-1}[d_{t+i}] &= \alpha_1 + \alpha_2 E_{t-1}[d_{t+i-1}].
\end{aligned}
\tag{4.8}
$$

In a world where all agents have rational and homogeneous beliefs about the future (i.e. a rational representative-agent model) the asset price should reflect the expected fundamental

value of the asset:

$$p_t = p_t^* + \eta_t \quad \eta_t \sim t_3(0, \sigma_\eta^2). \tag{4.9}$$

The *i.i.d.* error term $\eta_t$ is added so that $p_t$ is not an exact linear function of $d_{t-1}$, which would make the parameters in a VAR including prices and dividends not identifiable. The error term can however be justified as noise due to trading frictions. As discussed in Section 4.2, the error terms are drawn from a t-distribution. This is for empirical rather than theoretical considerations. Even though Kasa et al. (2010) address heterogeneous beliefs and nonfundamentalness in a theoretical context with a linear Gaussian model, non-Gaussian data are required for empirical detection of noncausality.

A more general version of model (4.8)-(4.9) relaxes the assumptions of homogeneity and rationality and allows for heterogeneous beliefs. I follow the asset-pricing model proposed by Brock and Hommes (1998), featuring many types of boundedly rational agents who form different beliefs about the future. With $H$ different types of agents, asset prices are determined by the following equation:

$$p_t \quad = \quad \sum_{h=1}^{H} \frac{n_{h,t} E_{h,t-1}[p_{t+1} + d_{t+1}]}{1+r} + \eta_t, \tag{4.10}$$

where $E_{h,t}(\cdot)$ represents the expectation formation mechanism of agent type $h$ and $n_{h,t}$ is the fraction of the population behaving according to type $h$ at time $t$. In the special case that $H = 1$ and $E_{1,t}(\cdot)$ denotes rational expectations $E_t(\cdot)$, (4.10) reduces to (4.9). To introduce heterogeneous beliefs it is useful to formulate (4.10) in deviation from the fundamental value:

$$x_t \quad = \quad \sum_{h=1}^{H} \frac{n_{h,t} f_{h,t}}{1+r} + \eta_t, \tag{4.11}$$

with $x_t = p_t - p_t^*$ is the realized difference from the fundamental value and $f_{h,t} = E_{h,t-1}[p_{t+1}] - E_{t-1}[p_{t+1}^*]$. Following Brock and Hommes (1998), agents hold identical beliefs about the fundamental value, but disagree on the dynamics of the deviation from the fundamental value. In particular, each type applies linear prediction rules based on lagged prices to form their

expectations:

$$f_{h,t} = g_h x_{t-1} + b_h. \tag{4.12}$$

The fraction of each type, $n_{h,t}$, varies over time according to evolutionary dynamics. The type of agent that realizes a high profit from trading in the previous period will become more influential in the next period:

$$n_{h,t} = \frac{\exp(\beta U_{h,t-1})}{\sum\limits_{i=1}^{H} \exp(\beta U_{i,t-1})}, \tag{4.13}$$

where $U_{h,t} = (x_t - (1+r)x_{t-1})(f_{h,t-1} - (1+r)x_{t-2}) - c_h$ denote the realized profits for each type, such that the fractions of all types add up to one. A full derivation of these equations is provided by Brock and Hommes (1998). These evolutionary dynamics are comparable to the 'forecasting the forecasts of others' property considered by Townsend (1983) and Kasa et al. (2010): Agents do not commit only to their own beliefs, but take into consideration the expectations of other agents, knowing that the expectations of others have a direct effect on asset prices. The parameter $\beta$ defines the willingness or capability of agents to switch to another strategy.

I now consider an example with two different agent types ($H = 2$): Optimists and pessimists (or bulls and bears). The optimist type forms expectations with a positive bias, while the pessimist type forms expectations with a negative bias:

$$
\begin{aligned}
f_{O,t} &= b \\
f_{P,t} &= -b,
\end{aligned}
\tag{4.14}
$$

with $b \geq 0$. This model reduces to the representative-agent benchmark (4.9) if $b = 0$. Optimists believe the asset is undervalued while pessimists believe the asset is overvalued. This disagreement could be the result of heterogeneous information on the fundamentals: The optimists (pessimist) receives positive (negative) signals about future fundamentals, although also other factors such as different levels of risk-aversion could cause the different beliefs.

Another, widely cited, example of the model by Brock and Hommes (1998) features fundamentalists and chartists. The fundamentalist believes deviations from the fundamental value should disappear:

$$f_{F,t} = 0. \tag{4.15}$$

The other type is the chartist or trend-follower, who believes deviations from the fundamental value in the previous period will persist:

$$f_{C,t} = g_C x_{t-1}. \tag{4.16}$$

The parameter $g_C$ defines the difference between the behavior of the agents. When $g_C = 0$, both types are identical. When $0 < g_C < 1 + r$, both types agree that deviations from the fundamental value should disappear over time, but they disagree about the pace of this correction. In Brock and Hommes (1998) $g_C \geq 1 + r$, meaning the chartists believe that the asset price will diverge from the fundamental value. Fundamentalists will therefore buy stocks when the price is under its fundamental valuation and sell when it is above. Chartists act the other way around which may create both positive and negative stock price bubbles even in the absence of random shocks (Brock and Hommes, 1998). Chartists are commonly thought of as technical traders, although Parke and Waters (2007) argue that similar behavior could be observed when agents experiment with different information sets to form expectations. The model with fundamentalists and chartists reduces to the representative-agent benchmark (4.9) when $g_C = 0$, or $n_{F,t} = 1 \; \forall t$.

I simulate dividends (4.7) and asset prices according to the representative-agent model (4.9), the optimist-pessimist model (4.10)-(4.14) and the fundamentalist-chartist model (4.10)-(4.13) and (4.15)-(4.16). Plots of 200 simulated observations of the asset prices under each model are given in Figure 4.1, together with the calibration of the parameters. The calibration of the profit functions and switching probabilities (4.13) is identical to the calibration by Brock and Hommes (1998). Figure 4.1 shows that under the representative-agent model, the difference between the fundamental values and the realized price is *i.i.d.* random noise (top panel). With the fundamentalist-chartist model, longer lasting deviations are observed. Thinking of annual
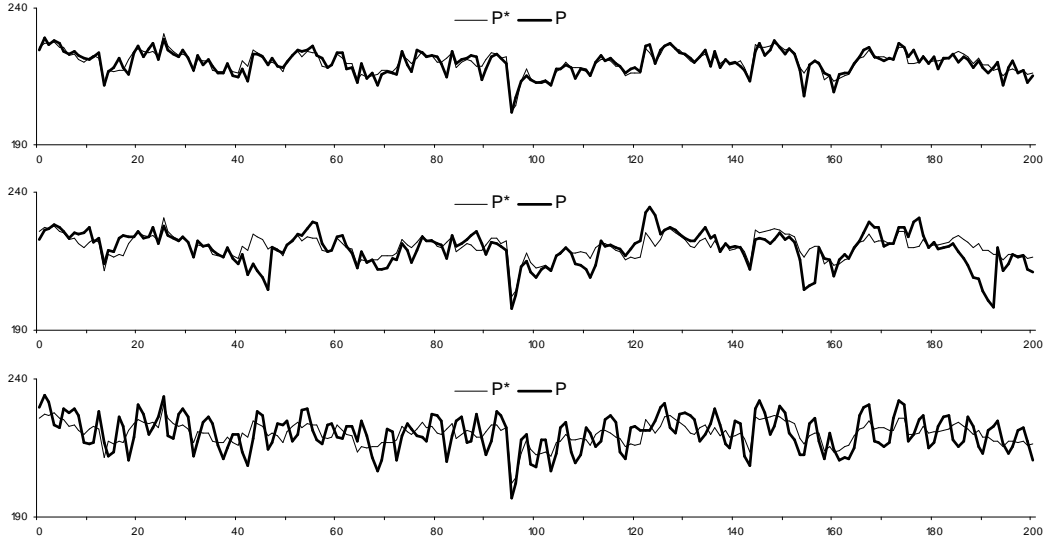
Figure 4.1: Simulated asset prices. Fundamental values and realized prices generated by: Representative agent *(Top panel)*, Fundamentalists and Chartists *(Middle panel)* and Optimists and Pessimists *(Bottom panel)*. *Calibration:* $\alpha_1 = 4$, $\alpha_2 = 0.8$, $\sigma_\varepsilon^2 = 1$, $r = 0.1$, $\sigma_\eta^2 = 2$, $\beta = 3.6$, $g_C = 1.2$, $c_F = 1$, $c_C = c_O = c_P = 0$, $b = 5.5$

data, the middle panel shows several examples of stock price bubbles lasting up to a decade. Finally, the bottom panel of Figure 4.1 shows the optimist-pessimist model, with continuous cycles of overvaluation followed by undervaluation lasting just a couple of years.

Apart from the calibration mentioned in Figure 4.1, the models are simulated with five different values values for $b$ and $g_C$, measuring the discrepancy between beliefs of optimists and pessimists and of chartists and fundamentalists respectively. The bias parameter $b$ is calibrated 1.1, 2.2, 3.3, 4.4 and 5.5, corresponding to a discrepancy between optimists' and pessimists' beliefs equal to respectively 1, 2, 3, 4 and 5% of the average fundamental value. The parameter $g_C$ is calibrated at 0.8, 0.9, 1.0, 1.1 and 1.2. Larger values of $g_C$ are not possible, as this model becomes unstable and converges to infinity when $g_C \geq (1+r)^2$ (Brock and Hommes, 1998).

After each simulation, the model selection procedure described in Section 4.2 is applied to determine whether the VAR including (demeaned) prices and dividends $(p_t, d_t)'$ is causal or noncausal. Since dividends follow a stationary AR process, there is no need to take (log) differences. This process is repeated 5000 times. Table 4.4 shows how often causal and noncausal specifications are selected for each model.

81

**TABLE 4.4**

| Representative agent | | | | | |
|---|---|---|---|---|---|
| Causal | 98% | | | | |
| Noncausal | 2% | | | | |

| *Optimists and Pessimists* | | | | | |
|---|---|---|---|---|---|
| $b$ | 1.1 | 2.2 | 3.3 | 4.4 | 5.5 |
| Causal | 98% | 78% | 66% | 63% | 60% |
| Noncausal | 2% | 22% | 34% | 37% | 40% |

| *Fundamentalists and Chartists* | | | | | |
|---|---|---|---|---|---|
| $g_C$ | 0.8 | 0.9 | 1.0 | 1.1 | 1.2 |
| Causal | 92% | 83% | 62% | 33% | 10% |
| Noncausal | 8% | 17% | 38% | 67% | 90% |

*Notes:* Percentage of causal and noncausal outcomes of the VAR for $(p_t, d_t)'$ after 5000 simulations of a representative-agent model (4.9) and of two heterogeneous-agents models (4.10)-(4.16) at multiple calibrations. The representative-agent model is identical to the two heterogeneous-agents models when $b = g_C = 0$ The sample size in each simulation is 200 observations.

With a representative agent the VARs of prices and dividends are found to be almost exclusively causal. However, with heterogeneous agents noncausality is found more often, up to 40% of the simulations with the optimist-pessimist model and even up to 90% with the fundamentalist-chartist model, even though all types of agents considered are fully backward-looking in the sense that they base their decisions only on past prices and dividends. Moreover, Table 4.4 clearly shows that noncausality is selected more often when the discrepancy between agents' beliefs (measured by $b$ and $g_C$) increases. These results confirm that heterogeneous beliefs are a potential source of noncausality. This is consistent with the simulation results in Section 4.3, since the fractions and strategies of each type of agent are unobservable and therefore omitted from the estimated model. Parke and Waters (2007) note that asset prices are generated by a process $P_t = f(\Omega_{t-1}, n_t, \varepsilon_t)$, where $\Omega_{t-1}$ includes all past prices and dividends and $n_t$ include the fractions of each type. In this case an econometrician will have access to $\Omega_{t-1}$, but can not observe behavior or expectations. An estimated model will therefore be of the form $P_t = \hat{f}(\Omega_{t-1}, \hat{\varepsilon}_t)$, so that $n_t$ is an omitted variable.

## 4.5 Conclusion

This paper presents empirical results confirming that, within the context of linear (vector) autoregressions, asset prices show a dependence on future observations and are therefore noncausal. A simulation study shows that the existence of heterogeneous beliefs is a potential source of noncausality. In this example, the econometrician has a smaller information set available than the actual agents in the economy and therefore misspecifies the agents' expectations formation mechanism. When only realized market data are observed, an important piece of information about the asset pricing process is omitted, namely the expectations and fractions of each type of agent.

Investor heterogeneity is not the only potential source of noncausality. Also in a representative agent model, the evolution of asset prices may depend on unobservable elements such as a time-varying (stochastic) discount factor.

The result that asset prices are noncausal, raises opportunities for further research. Noncausal forecasting methods proposed by Lanne et al. (2012 a,b) may be helpful in predicting asset prices and returns. Moreover, in structural modeling of asset price dynamics, the issue of nonfundamentalness should be addressed (e.g. Forni et al. 2009, Fernandez-Villaverde et al. 2007).

## References

Alessi, L., M. Barigozzi, and M. Capasso: 2011, 'Nonfundamentalness in Structural Econometric Models: A Review'. *International Statistical Review* **79**(1), 16–47.

Boswijk, H. P., C. H. Hommes, and S. Manzan: 2007, 'Behavioral heterogeneity in stock prices'. *Journal of Economic Dynamics and Control* **31**(6), 1938–1970.

Breidt, F. J., R. A. Davis, K.-S. Lh, and M. Rosenblatt: 1991, 'Maximum likelihood estimation for noncausal autoregressive processes'. *Journal of Multivariate Analysis* **36**(2), 175–98.

Brock, W. A. and C. H. Hommes: 1998, 'Heterogeneous beliefs and routes to chaos in a simple asset pricing model'. *Journal of Economic Dynamics and Control* **22**(8-9), 1235–1274.

Brockwell, P. J. and R. A. Davis: 1991, *Time Series: Theory and Methods, Second Edition*. New York, NY: Springer-Verlag, 1991 edition.

Campbell, J. Y. and R. J. Shiller: 1987, 'Cointegration and Tests of Present Value Models'. *Journal of Political Economy* **95**(5), 1062–88.

Campbell, J. Y. and R. J. Shiller: 1988, 'The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors'. *Review of Financial Studies* **1**(3), 195–228.

Fernandez-Villaverde, J., J. F. Rubio-Ramirez, T. J. Sargent, and M. W. Watson: 2007, 'ABCs (and Ds) of Understanding VARs'. *American Economic Review* **97**(3), 1021–1026.

Forni, M., D. Giannone, M. Lippi, and L. Reichlin: 2009, 'Opening The Black Box: Structural Factor Models With Large Cross Sections'. *Econometric Theory* **25**(05), 1319–1347.

Hansen, L. P. and T. J. Sargent: 1991, 'Two Difficulties in Interpreting Vector Autoregressions'. In: L. P. Hansen and T. J. Sargent (eds.): *Rational Expectations Econometrics*. Westview Press, Inc., Boulder, CO, pp. 77–119.

Kasa, K., T. B. Walker, and C. H. Whiteman: 2010, 'Heterogeneous Beliefs and Tests of Present Value Models'. Unpublished manuscript.

Lanne, M., A. Luoma, and J. Luoto: 2012a, 'Bayesian Model Selection and Forecasting in Noncausal Autoregressive Models'. *Journal of Applied Econometrics* **27**(5), 812–830.

Lanne, M., J. Luoto, and P. Saikkonen: 2012b, 'Optimal Forecasting of Noncausal Autoregressive Time Series'. *International Journal of Forecasting* **28**(3), 623 – 631.

Lanne, M. and P. Saikkonen: 2011a, 'GMM Estimation with Noncausal Instruments'. *Oxford Bulletin of Economics and Statistics* **73**(5), 581–592.

Lanne, M. and P. Saikkonen: 2011b, 'Noncausal Autoregressions for Economic Time Series'. *Journal of Time Series Econometrics* **3**(3), Article 2.

Lanne, M. and P. Saikkonen: 2013, 'Noncausal vector autoregression'. *Econometric Theory* (forthcoming).

Lof, M.: 2013, 'Noncausality and Asset Pricing'. *Studies in Nonlinear Dynamics and Econometrics* **17**(2), 211–220.

Meitz, M. and P. Saikkonen: 2013, 'Maximum likelihood estimation of a noninvertible ARMA model with autoregressive conditional heteroskedasticity'. *Journal of Multivariate Analysis* **114**, 227 – 255.

Parke, W. R. and G. A. Waters: 2007, 'An evolutionary game theory explanation of ARCH effects'. *Journal of Economic Dynamics and Control* **31**(7), 2234–2262.

Shiller, R. J.: 2005, *Irrational exuberance*. Princeton University Press.

Townsend, R. M.: 1983, 'Forecasting the Forecasts of Others'. *Journal of Political Economy* **91**(4), 546–88.

# Chapter 5

## GMM estimation with noncausal instruments under rational expectations[1]

## 5.1 Introduction

In a recent paper, Lanne and Saikkonen (2011a) warn against the use of the generalized method of moments (GMM; Hansen, 1982), when the instruments are lags of variables that admit a noncausal autoregressive representation. With such noncausal instruments, the two-stage least squares (2SLS) estimator is shown to be inconsistent under certain assumptions on the distribution of the error term in the regression model. In this paper, I make no explicit assumptions on this distribution. Instead, the errors are implied by a rational expectations equilibrium and are in fact prediction errors. GMM estimation is in this case consistent even when the instruments are noncausal.

The application of GMM is widespread in empirical macroeconomics and finance (see, e.g. the survey by Hansen and West, 2002). Typical examples include the estimation of an Euler equation (e.g. Hansen and Singleton, 1982, Campbell and Mankiw, 1990) or a Philips curve (e.g. Gali and Gertler, 1999). In these examples, the moment conditions are based on the assumption of rational expectations, implying that error terms must be orthogonal to all observed information. A lagged value of any observable variable should therefore be a valid instrument.

---

[1]This chapter is based on an article forthcoming in the *Oxford Bulletin of Economics and Statistics* (Lof, 2013)

Lanne and Saikkonen (2011a) consider a linear regression model with a single regressor:

$$y_t = \delta x_t + \eta_t, \tag{5.1}$$

and evaluate the situation in which $x_t$ is noncausal. A variable is noncausal, when it follows a noncausal autoregressive process, that allows for dependence on both leading and lagging observations. A noncausal AR$(r,s)$ process, as defined by Lanne and Saikkonen (2011b), depends on $r$ past and $s$ future observations:

$$\phi(L)\varphi(L^{-1})x_t = \varepsilon_t, \tag{5.2}$$

with $\phi(L) = 1 - \phi_1 L - ... - \phi_r L^r$, $\varphi(L^{-1}) = 1 - \varphi_1 L^{-1} - ... - \varphi_r L^{-s}$, $\varepsilon_t \sim i.i.d.(0,\sigma^2)$ and $L$ is a standard lag operator ($L^k y_t = y_{t-k}$). A noncausal AR process has an infinite-order moving average (MA) representation that is both backward- and forward-looking:

$$x_t = \varphi(L^{-1})^{-1}\phi(L)^{-1}\varepsilon_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \tag{5.3}$$

in which $\psi_j$ is the coefficient of $z^j$ in the Laurent-series expansion of $\varphi(z^{-1})^{-1}\phi(z)^{-1}$ (Lanne and Saikkonen, 2011b). When $x_t$ is a vector, (5.2) defines a noncausal VAR$(r,s)$ process (Lanne and Saikkonen, 2013).

Lanne and Saikkonen (2011a) make the following distributional assumption on the errors in (5.1) and (5.2):

$$(\varepsilon_t, \eta_t)' \sim i.i.d.(0,\Omega), \tag{5.4}$$

with nonzero covariance: $\Omega_{12} = E[\varepsilon_t \eta_t] \neq 0$. Since $x_t$ and $\eta_t$ are correlated, OLS estimation of equation (5.1) is inconsistent. However, the MA representation (5.3) reveals that also 2SLS estimation is inconsistent when lags of $x_t$ are used as instruments, since these lags depends on $\varepsilon_t$ and are therefore correlated with $\eta_t$: $E[x_{t-i}\eta_t] = \psi_{-i}E[\varepsilon_t \eta_t] = \psi_{-i}\Omega_{12}$, which is nonzero if $\varphi_j \neq 0$, for some $j \in \{1,..,s\}$ in equation (5.2). The next section shows that this inconsistency does not hold under the assumption of rational expectations.

## 5.2 Prediction errors

For ease of exposition, I consider the linear regression model (5.1), with $x_t$ generated by a Gaussian first-order noncausal autoregression. Lof (2011) provides additional simulation results showing robustness to non-Gaussian and higher-order autoregressive specifications of $x_t$. The result is further illustrated in the next section with a nonlinear asset pricing model.

If the dependent variable $y_t$ in the linear regression (5.1) is the outcome of a rational expectations equilibrium, the error term error term $\eta_t$ has the interpretation of a prediction error:

$$
\begin{aligned}
y_t &= \delta E_{t-1}[x_t] \\
\eta_t &= -\delta (x_t - E_{t-1}[x_t]),
\end{aligned}
\tag{5.5}
$$

in which $E_{t-1}[\cdot] \equiv E[\cdot \mid \Theta_{t-1}]$ and $\Theta_{t-1}$ denotes the information set which includes all information observable in period $t-1$. In this case, all variables belonging to $\Theta_{t-1}$ are uncorrelated with $\eta_t$. Lagged values of $x_t$, assuming they are observable ($x_{t-i} \in \Theta_{t-1}$, $i \geq 1$), are therefore valid instruments regardless of their dynamic properties:

$$
\begin{aligned}
E[x_{t-i}\eta_t] &= E[x_{t-i}E_{t-1}[\eta_t]] && \{i \geq 1\} \\
&= E[x_{t-i}E_{t-1}[-\delta(x_t - E_{t-1}[x_t])]] \\
&= -\delta E[x_{t-i}(E_{t-1}[x_t] - E_{t-1}[x_t])] = 0.
\end{aligned}
\tag{5.6}
$$

To see how this differs from the result by Lanne and Saikkonen (2011a), assume the regressor $x_t$ to be generated by a Gaussian first-order noncausal autoregressive process, AR(0,1):

$$
\begin{aligned}
x_t &= \alpha x_{t+1} + \varepsilon_t \\
&= \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t+j},
\end{aligned}
\tag{5.7}
$$

with $\varepsilon_t \sim N(0, \sigma^2)$. Since $x_t$ is Gaussian, the noncausal process (5.7) is indistinguishable from a causal AR(1,0) process, and its optimal forecast is identical to the causal case: $E_{t-1}[x_t] = \alpha x_{t-1}$ (Lanne et al., 2012). The realized prediction error (assuming the true value of $\alpha$ is known) is

then:

$$
\begin{aligned}
e_t &= x_t - E_{t-1}[x_t] \\
&= x_t - \alpha x_{t-1}
\end{aligned}
\tag{5.8}
$$

The prediction error $e_t$ is the true 'innovation' in $x_t$ and is, other than in a causal autoregression, not equal to the error term $\varepsilon_t$. In fact, from the MA representation of $x_t$ (5.7), it is straightforward to see that the prediction error is correlated with lags and leads of $\varepsilon_t$:

$$
\begin{aligned}
E[e_t \varepsilon_{t-i}] &= E[x_t \varepsilon_{t-i}] - \alpha E[x_{t-1} \varepsilon_{t-i}] \\
&= \begin{cases}
0 - \alpha \sigma^2 &= -\alpha \sigma^2 & \{i = 1\} \\
\alpha^i \sigma^2 - \alpha \alpha^{i+1} \sigma^2 &= (1 - \alpha^2) \alpha^i \sigma^2 & \{i < 1\} \\
0 - 0 &= 0 & \{i > 1\},
\end{cases}
\end{aligned}
\tag{5.9}
$$

Since the implied error term $\eta_t$ is an exact linear function of the prediction error $e_t$ ($\eta_t = -\delta e_t$), $\eta_t$ is correlated with leads and lags of $\varepsilon_t$, which contradicts the assumption (5.4) made by Lanne and Saikkonen (2011a). The prediction errors $e_t$ and $\eta_t$ are, however, uncorrelated with lags of $x_t$:

$$
\begin{aligned}
E[e_t x_{t-i}] &= E[x_t x_{t-i}] - \alpha E[x_{t-1} x_{t-i}] \\
&= \alpha^i E[x_t^2] - \alpha \alpha^{i-1} E[x_t^2] = 0 \quad \{i \geq 1\},
\end{aligned}
\tag{5.10}
$$

which means that lags of $x_t$ are valid instruments for estimating (5.1), regardless of whether $x_t$ is causal or noncausal.

This result can be extended to a multivariate context. Let $\mathbf{x}_t$ be a $K$-dimensional vector of variables that is generated by a noncausal VAR$(0, 1)$ process:

$$
\mathbf{x}_t = B\mathbf{x}_{t+1} + \varepsilon_t,
\tag{5.11}
$$

with $\varepsilon_t \sim N(\mathbf{0}, \Sigma_B)$, while $\mathbf{x}_t^*$ follows a causal VAR$(1, 0)$ process:

$$
\mathbf{x}_t^* = A\mathbf{x}_{t-1}^* + \varepsilon_t^*,
\tag{5.12}
$$

with $\varepsilon_t^* \sim N(\mathbf{0}, \Sigma_A)$. The processes $\mathbf{x}_t$ and $\mathbf{x}_t^*$ are identical in first- and second-order moments when:

$$
\begin{aligned}
B &= \Gamma_0^* A' \Gamma_0^{-1} \\
\Sigma_B &= \Gamma_0^* - B\Gamma_0 B',
\end{aligned}
\tag{5.13}
$$

in which the covariance functions are defined by:

$$
\begin{aligned}
\Gamma_0 &= E\left[\mathbf{x}_t \mathbf{x}_t'\right] &= B\Gamma_0 B' + \Sigma_B \\
\Gamma_0^* &= E\left[\mathbf{x}_t^* \mathbf{x}_t^{*\prime}\right] &= A\Gamma_0^* A' + \Sigma_A.
\end{aligned}
\tag{5.14}
$$

It is straightforward to verify that $\Gamma_0 = \Gamma_0^*$, when (5.13) holds. Under these conditions, also the autocovariance functions of $\mathbf{x}_t$ and $\mathbf{x}_t^*$ are identical:

$$
\begin{aligned}
\Gamma_{-i} &= E\left[\mathbf{x}_t \mathbf{x}_{t+i}'\right] &= B^i \Gamma_0 \\
\Gamma_i^* &= E\left[\mathbf{x}_t^* \mathbf{x}_{t-i}^{*\prime}\right] &= A^i \Gamma_0^*.
\end{aligned}
\tag{5.15}
$$

Since $\Gamma_{-i} = \Gamma_i'$, the autocovariance function of the causal and noncausal processes are identical if and only if $B^i \Gamma_0 = \Gamma_0^* A'^i$, or equivalently: $B^i = \Gamma_0^* A'^i \Gamma_0^{-1}$, which is satisfied for all $i$ when $B = \Gamma_0^* A' \Gamma_0^{-1}$ and $\Gamma_0 = \Gamma_0^*$.

The equivalence in first- and second-order moments implies that, under Gaussianity, the processes (5.11) and (5.12) are indistinguishable, so $E_{t-1}[\mathbf{x}_t] = A\mathbf{x}_{t-1}$ is the optimal forecast for both the causal and noncausal process (Lanne et al., 2012). The vector of forecast errors is then, analogous to equation (5.8), $\mathbf{e}_t = \mathbf{x}_t - A\mathbf{x}_{t-1}$. As in the univariate case (5.9)-(5.10) $\mathbf{e}_t$ is correlated with lags and leads of $\varepsilon_t$, but uncorrelated with lags of $\mathbf{x}_t$:

$$
\begin{aligned}
E\left[\mathbf{e}_t \mathbf{x}_{t-i}'\right] &= \Gamma_{-i}' - A^i \Gamma_0 \\
&= \Gamma_0 B'^i - \Gamma_0 B'^i \Gamma_0^{-1} \Gamma_0 = 0 \quad \{i \geq 1\}.
\end{aligned}
\tag{5.16}
$$

Under the assumption that the error term in a regression equation like (5.1) is a linear combination of prediction errors: $\eta_t = \gamma' \mathbf{e_t}$, lags of $\mathbf{x}_t$ are uncorrelated with this error term ($E\left[\eta_t \mathbf{x}_{t-i}\right] = 0$ $\forall i \geq 1$) and are therefore valid instruments.

## 5.3 Example: Consumption-based asset pricing

Consumption-based asset pricing was amongst the first applications of GMM (Hansen and Singleton, 1982). The model to estimate is an Euler equation relating financial returns ($R_t = P_{t-1}^{-1}(P_t + D_t)$) to the marginal rate of substitution:

$$E_{t-1}\left[\beta \frac{u'(C_t)}{u'(C_{t-1})} R_t\right] = 1, \tag{5.17}$$

in which $P_t$ refers to asset prices, $D_t$ to dividends and $C_t$ to consumption. Multiplying this optimality condition with a vector of predetermined instruments $\mathbf{z}_{t-1}$ and assuming a constant relative-risk aversion utility function ($u(C_t) = (1 - \gamma)^{-1} C_t^{1-\gamma}$) gives the required moment conditions for GMM estimation:

$$E\left[\left(\beta \left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} R_t - 1\right) \mathbf{z}_{t-1}\right] = 0. \tag{5.18}$$

This approach has become leading practice in empirical finance (see e.g. Ludvigson, 2011, for a recent survey). It is illustrative to see that a simple regression model, similar to (5.1), is obtained after log-linearizing the Euler equation:

$$r_t = \mu + \gamma \triangle c_t + \eta_t, \tag{5.19}$$

in which $r_t = \log(R_t)$ and $c_t = \log(C_t)$. Yogo (2004) shows that the error term $\eta_t$ is in this case indeed a linear combination of prediction errors, as assumed in Section 5.2:

$$\eta_t = (r_t - E_{t-1}[r_t]) - \gamma(\triangle c_t - E_{t-1}[\triangle c_t]), \tag{5.20}$$

I simulate returns and consumption according to (5.17), to verify that the GMM estimator is consistent even if the instruments are noncausal. The first step is to define log consumption and dividend growth as a first-order VAR process, $(\triangle c_t, \triangle d_t)' = \mathbf{x}_t$, in which $d_t = \log(D_t)$. This process may be causal or noncausal, i.e. is generated by equation (5.12) or (5.11). The restric-

**TABLE 5.1:** Calibration

| | | $A$ | | $\Sigma_A$ | | $\beta$ | $\gamma$ |
|---|---|---|---|---|---|---|---|
| (i) | $(\triangle c_t, \triangle d_t)' \equiv \mathbf{x}_t$ | $\begin{pmatrix} -0.161 & 0.017 \\ 0.414 & 0.117 \end{pmatrix}$ | | $\begin{pmatrix} 0.0012 & 0.0018 \\ 0.0018 & 0.014 \end{pmatrix}$ | | 0.97 | 1.3 |
| (ii) | $\triangle c_t = \triangle d_t \equiv x_t$ | $-0.14$ | | 0.009 | | 0.97 | 1.3 |

*Notes:* Calibrations of $A$, $\Sigma_A$, $\beta$ and $\gamma$ in the Euler equation (5.17). The first calibration (*i*) follows Wright (2003). In the second calibration (*ii*), consumption and dividends are identical as in a Lucas-tree economy (Lucas, 1978). The autoregressive process may be causal or noncausal. The parameter values of the noncausal autoregressive process are derived from $A$ and $\Sigma_A$ according to equation (5.13)

tions (5.13) apply, so both specifications are identical in their mean, variance and autocorrelation function. Given a simulated sample of consumption and dividends, I generate returns following the approach of Tauchen and Hussey (1991). Multiplying equation (5.17) by $\frac{P_{t-1}}{D_{t-1}}$, results in a nonlinear stochastic difference equation describing the dynamics of the price-dividend (PD) ratio:

$$\frac{P_{t-1}}{D_{t-1}} = E_{t-1}\left[\beta\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}\frac{D_t}{D_{t-1}}\left(1 + \frac{P_t}{D_t}\right)\right], \tag{5.21}$$

which can be simulated by calibrating a discrete-valued Markov chain that approximates the conditional distribution of consumption and dividend growth. Details on this approximation for the causal VAR are provided by Tauchen (1986) and this method can be implemented for the noncausal VAR too, as the conditional distributions of the causal and noncausal processes are identical under Gaussianity and the restrictions in (5.13). Returns are then computed from the simulated dividends and PD ratios.

I consider two different calibrations of the matrices $A$ and $\Sigma_A$ in (5.12), which are given in Table 5.1. The first calibration (*i*) of $A$ and $\Sigma_A$ is following Wright (2003) and is based on actual data on annual consumption and dividend growth. In the second example (*ii*), consumption growth follows a univariate AR(1,0) or AR(0,1) process, which is calibrated to have identical variance and autocorrelation as consumption growth in the first calibration, while dividend growth is set equal to consumption growth. This is an example of a "Lucas-tree economy", in which household income consists of dividends alone. It is well known that in this case there exists a no-trade equilibrium in which households consume their entire endowment of dividends (Lucas, 1978).

I use the simulated returns and consumption growth rates to estimate $\beta$ and $\gamma$ by two-step

efficient GMM, based on the moment conditions (5.18), using $\mathbf{z}_{t-1} = \left(1, \frac{C_{t-1}}{C_{t-2}}, R_{t-1}\right)'$ as instruments, following Hansen and Singleton (1982). I consider 10,000 replications with sample sizes of 50 and 1000 observations.

Table 5.2 displays the simulation results. The main result is that for both calibrations, non-causality of the instruments seems to have no effect on the finite-sample or asymptotic properties of the GMM estimator. In both cases, the GMM estimates of $\beta$ and $\gamma$ are rather poor for small samples, but improve for larger samples. It is clear that the inconsistency of the estimator derived by Lanne and Saikkonen (2011a), does not hold under the assumptions in this model.

Figure 5.1 shows plots of the correlation between the Euler-equation errors $u_t = \widehat{\beta}\left(\frac{C_t}{C_{t-1}}\right)^{-\widehat{\gamma}} R_t - 1$ and lags and leads of $\varepsilon_t$ and $\frac{C_t}{C_{t-1}}$. These correlation plots are consistent with the results derived in Section 5.2: When consumption is generated by a causal process, $u_t$ is only correlated with $\varepsilon_t$, but not with its leads and lags. With noncausal consumption, on the other hand, the error term $u_t$ is correlated with lags and leads of $\varepsilon_t$, so assumption (5.4) does not hold. Despite these intertemporal correlations, the important point to notice is that lags of $\frac{C_t}{C_{t-1}}$ are uncorrelated with $u_t$, which means they are valid instruments.

**TABLE 5.2:** Simulation results

| Calibration | Causal | | | | Noncausal | | | |
|---|---|---|---|---|---|---|---|---|
| | (i) | | (ii) | | (i) | | (ii) | |
| $T$ | 50 | 1000 | 50 | 1000 | 50 | 1000 | 50 | 1000 |
| $\beta$ | 0.965 | 0.970 | 0.970 | 0.970 | 0.965 | 0.970 | 0.970 | 0.970 |
| | (0.030) | (0.004) | (0.001) | (0.000) | (0.030) | (0.004) | (0.001) | (0.000) |
| $\gamma$ | 1.742 | 1.293 | 1.115 | 1.285 | 1.743 | 1.292 | 1.114 | 1.285 |
| | (3.556) | (0.810) | (0.202) | (0.067) | (3.580) | (0.809) | (0.190) | (0.067) |

*Notes:* Average two-step efficient GMM estimates and standard deviations (in parenthesis) of $\beta$ and $\gamma$, model (5.17), after 10,000 replications of sample size $T$. Instruments are $\mathbf{z}_{t-1} = \left(1, \frac{C_{t-1}}{C_{t-2}}, R_{t-1}\right)'$. Consumption and dividends are generated by a causal or noncausal autoregressive process. Returns are computed following the approach of Tauchen and Hussey (1991). Calibrations of the Euler equation and autoregressive processes are given in Table 5.1.
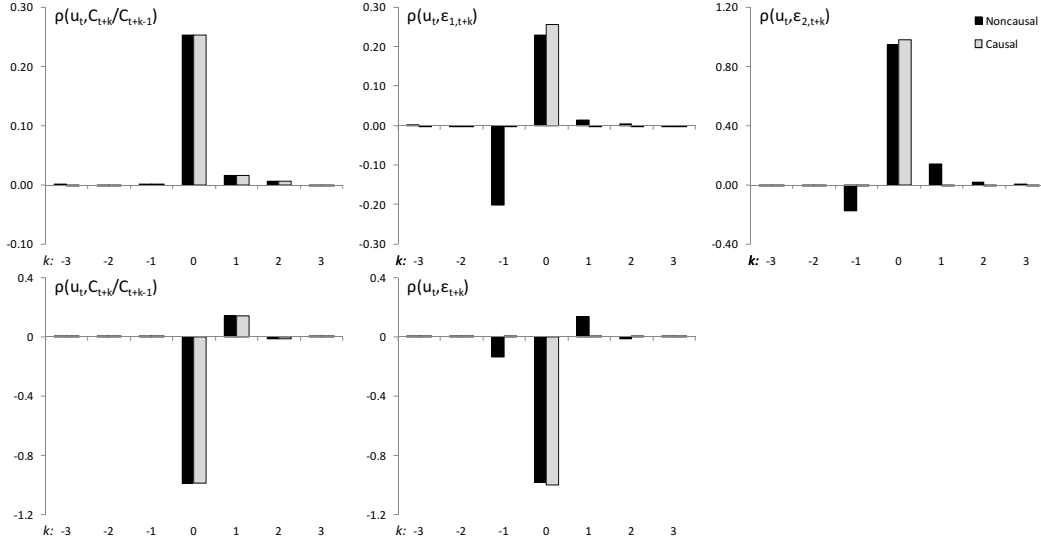
Figure 5.1: Correlations of errors and instruments. Correlations between residuals from GMM estimates in Table 5.2: $u_t = \widehat{\beta} \left( \dfrac{C_t}{C_{t-1}} \right)^{-\widehat{\gamma}} R_t - 1$ and lags and leads of $\varepsilon_t$ and $\dfrac{C_t}{C_{t-1}}$, for calibration (*i*), top, and (*ii*), bottom.

## 5.4   Conclusion

Instead of making explicit distributional assumptions on the error terms in a regression model, I argue that these errors are to be interpreted as prediction errors. This interpretation is consistent with the approach by Hansen and Singleton (1982), amongst others, who base GMM estimation on moment conditions implied by rational-expectations theories. All variables included in the information set on which agents condition to form expectations are in this case valid instruments, whether these are causal or noncausal. This is good news to those who apply GMM, although other caveats, such as weak instruments or misspecified economic theories, are of course still around to complicate the tasks of applied econometricians.

## References

Campbell, J. Y. and N. G. Mankiw: 1990, 'Permanent Income, Current Income, and Consumption'. *Journal of Business & Economic Statistics* **8**(3), 265–79.

Gali, J. and M. Gertler: 1999, 'Inflation dynamics: A structural econometric analysis'. *Journal of Monetary Economics* **44**(2), 195–222.

Hansen, B. E. and K. D. West: 2002, 'Generalized Method of Moments and Macroeconomics'. *Journal of Business & Economic Statistics* **20**(4), 460–69.

Hansen, L. P.: 1982, 'Large Sample Properties of Generalized Method of Moments Estimators'. *Econometrica* **50**(4), 1029–54.

Hansen, L. P. and K. J. Singleton: 1982, 'Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models'. *Econometrica* **50**(5), 1269–86.

Lanne, M., J. Luoto, and P. Saikkonen: 2012, 'Optimal forecasting of noncausal autoregressive time series'. *International Journal of Forecasting* **28**(3), 623 – 631.

Lanne, M. and P. Saikkonen: 2011a, 'GMM Estimation with Noncausal Instruments'. *Oxford Bulletin of Economics and Statistics* **73**(5), 581–592.

Lanne, M. and P. Saikkonen: 2011b, 'Noncausal Autoregressions for Economic Time Series'. *Journal of Time Series Econometrics* **3**(3), Article 2.

Lanne, M. and P. Saikkonen: 2013, 'Noncausal vector autoregression'. *Econometric Theory* (forthcoming).

Lof, M.: 2011, 'GMM estimation with noncausal instruments under rational expectations'. *HECER Discussion Paper* (343).

Lof, M.: 2013, 'GMM estimation with noncausal instruments under rational expectations'. *Oxford Bulletin of Economics and Statistics* (forthcoming).

Lucas, R. E. J.: 1978, 'Asset Prices in an Exchange Economy'. *Econometrica* **46**(6), 1429–45.

Ludvigson, S. C.: 2011, 'Advances in Consumption-Based Asset Pricing: Empirical Tests'. *NBER Working Paper* (16810).

Tauchen, G.: 1986, 'Finite state markov-chain approximations to univariate and vector autoregressions'. *Economics Letters* **20**(2), 177–181.

Tauchen, G. and R. Hussey: 1991, 'Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models'. *Econometrica* **59**(2), 371–96.

Wright, J. H.: 2003, 'Detecting Lack of Identification in GMM'. *Econometric Theory* **19**(02), 322–330.

Yogo, M.: 2004, 'Estimating the Elasticity of Intertemporal Substitution when Instruments are Weak'. *The Review of Economics and Statistics* **86**(3), 797–810.