THE BONN JOURNAL OF ECONOMICS

A Student Economic Review

Volume I December 2012

Department of Economics University of Bonn

Legal Information

Publisher	The Bonn Journal of Economics Adenauerallee 24-42 53113 Bonn http://bje.uni-bonn.de bje@uni-bonn.de
V. i. S. d. P.	Prof. Dr. K. Sandmann Head of Department of Economics University of Bonn
Place of Publication	Bonn, Germany
ISSN	2195-3449

The Bonn Journal of Economics

Founded in 2012

VOLUME 1

Founded by

Johannes Hermle   Justus Inhoffen   Tobias Ruof

Department of Economics
University of Bonn

The Bonn Journal of Economics

Review of the Equity Premium Puzzle

Benjamin Große-Rüschkamp*

Meet the Equity Premium Puzzle

The equity premium, the excess return of equity over relatively risk-free government bonds is well-documented and is of the magnitude of $3-8$ percent per year, depending on the period considered.[1] One basic explanation is that the equity premium is a risk premium that investors receive for bearing the additional risk associated with holding equity instead of bonds. This is the notion borne by Capital Asset Pricing Models. Because of theoretical shortcomings, macroeconomists have developed a generalization of the traditional CAPM. Consumption-based asset pricing allows for richer microfoundations and also takes into account the intertemporal dimension of portfolio investing (Breeden (1991)). So due to its theoretical foundations, consumption-based asset pricing seems to offer an excellent point of departure to pose the question about the "right" size of the 'risk premium'. Using the predominant macroeconomic asset-pricing model, Mehra and Prescott (1985) found that only a small fraction of the premium could, in fact, be accounted for by risk aversion under the neoclassical representative-agent paradigm. The finding, termed 'Equity Premium Puzzle' withstood scrutiny of

* Benjamin Große-Rüschkamp received his degree in Economics (B. Sc.) from the University of Bonn in 2011. The present article refers to his bachelor thesis submitted in September 2011.

[1] Mehra and Prescott (2008a) report for the annual percentage equity premium: 3.62 (1889 – 1933); 8.07 (1934 – 2005); 7.48 (1946 – 2005); Own estimations give 7.94 (1934 – 2010).

the profession, and puzzled macroeconomists came up with a host of new frameworks and expansions on existing concepts to reconcile the data with theory.

Mehra and Prescott (1985) set out to empirically test the implications of Lucas' seminal paper (Lucas (1978)) in which he was first to rigorously spell out the idea of relating asset price behaviour to consumption in the context of a complete market general equilibrium economy. The core of the modelled economy forms the representative agent. Here, the assumption of complete markets and isoelastic utility are sufficient for aggregation because in equilibrium, the marginal utilities of all agents are proportional (see, for example, Ljungqvist and Sargent (2004)). By deciding on a consumption path $\{c\}_{t=0}^{\infty}$, the representative agent maximizes the sum of its discounted expected utilities, i.e. $\max E_0\{\sum_{t=0}^{\infty} \beta^t U(c_t)\}$. The utility function $U(c, \alpha)$ is defined as $U(c, \alpha) = \frac{c^{1-\alpha}-1}{1-\alpha}$. Here, α represents the coefficient of relative risk aversion; at the same time, the inverse α^{-1} represents the coefficient of the elasticity of intertemporal substitution. This measures the agent's willingness to substitute consumption intertemporally. As both coefficients are time-invariant (i.e. constant), this class of utility functions is also referred to as CRRA/CIES[2] or isoelastic utility. Conveniently, CRRA is invariant to scale transformations so the demand for assets with risky payoffs is a linear function of wealth. As this makes it independent of initial distribution of wealth or endowment, it allows for aggregation of utility and the use of a representative agent. However, it also implies that a change of the preferences that affect the attitude towards smoothing consumption across time also affects the attitude towards smoothing consumption across states. This assumption may be problematic and is revisited later on. In equilibrium, the following relationship on asset prices should hold: $P_t U'(c_t) = \beta E_t\{U'(c_{t+1})(P_{t+1} + d_{t+1})\}$. This stochastic Euler equation, as presented by Lucas (1978) (equation (6)) has a convenient

[2] CRRA: Constant Relative Risk Aversion; CIES: Constant Intertemporal Elasticity of Substition

interpretation. It expresses the intertemporal choice problem of the agent at time t. After performing some algebra we obtain what is known as 'Lucas asset pricing formula', namely $1 = E_t[m_{t+1}q_{t+1}]$, with $m_{t+1} = \beta\frac{u'(c_{t+1})}{u'(c_t)}$, where $q_{t+1} = 1 + R_{t+1}$ is the period gross return and m_{t+1} is a random variable and referred to as Stochastic Discount Factor. Performing some further operations, Mehra and Prescott calibrate the model. Using macroeconomic data, the free parameters remaining are β and α. Then, using values for parameters estimated by other economists,[3] Mehra and Prescott's 1985 estimation results in terms of admissible pairs of R^f and R^e that can be plotted using those values. Clearly off by more than an order of magnitude, the largest attainable value for the equity premium $R^e - R^f$ is 0.35, compared to 6.18 for the average estimated premium. On top, this value for the equity premium comes at the cost of a counterfactual average risk free rate of about 4 percent. This result represents the puzzle.

Preference-based solutions: Generalized Expected Utility and Habit Formation

The CRRA preference possesses many desirable traits, however, it links risk aversion and preference of consumption stability across time and across different states of nature. There is no compelling economic reason for why preferences concerning consumption patterns across time and states would have to be linked. It seems clear that increasing the risk-aversion of households by increasing the parameter value α of the original utility function will drive up the equity premium. Equity returns are more volatile, and thus perceived as more risky; so equity returns need to be high relative to risk-free returns to induce households to invest in them. This is even more true when increasing the value for α. So then,

[3] Friend and Blume (1975): $\alpha \geq 2$; Tobin and Dolde (1971) $\alpha = 1$; applying this information liberally, Mehra and Prescott (1985) restrict $\alpha \geq 10$.

why can we not just set α high and generate an equity premium that matches the observed? The issue is the following: by setting a high α, it is possible to produce a large enough equity premium through an increase of the household's risk aversion. At the same time, this also increases the household's aversion to unstable consumption paths over time. Given that consumption grows over time, there is little incentive to save. Rather, there is an urge to borrow against future income to increase consumption today. In the aggregate, however, households cannot borrow. What the inclination to borrow will do, though, is drive up returns of securities. Therefore, increasing α will raise returns; in particular the risk-less rate would be far higher than observed. Weil (1989) tests the asset-pricing implications in an economic environment otherwise identical to Mehra and Prescott (1985), but employing a GEU[4] preference ordering. A basic form of the aggregator used by Weil is the following: $U_t = \{c_t^{1-\rho} + \beta(E_t[U_{t+1}^{1-\gamma}])^{\frac{1-\rho}{1-\gamma}}\}^{\frac{1}{1-\rho}}$. This isoelastic utility function visibly disentangles the coefficient of risk aversion γ from the coefficient of intertemporal substitution, ρ. While intuition seems to suggest that a recalibration of preferences would significantly mitigate the Equity Premium Puzzle—adding a variable increases the degree of freedom for varying parameters—Weil (1989) demonstrates that this is not the case at all, and names the result of the counterfactual high rate, the 'Risk-free Rate Puzzle'. Disentangling the coefficients governing the attitude towards risk and intertemporal substitution does not remedy the familiar issues because the role of the coefficient of intertemporal substitution is actually strengthened by the separation: In a world of growing consumption, agents will once again have to be offered a counterfactually high return to be induced to postpone consumption.

Another potential way to solve the equity premium puzzle is to leave behind the assumption of per-period consumption as the only utility enhancing

[4] GEU: Generalized Expected Utility

measure, an approach grounded to some extend in the behavioural literature. An interesting case of attempts to resolve the puzzle is the use of models with habit formation: instantaneous utility is not simply derived from the per-period consumption-level, but is rather determined by relating it to the agent's history of consumption which may form an internal benchmark, also known as a habitual level of consumption ('internally measured habit'). In another variant, the agent derives utility from relative status, so his consumption is contrasted to the consumption-level of others ('externally measured habit'). In the first case, utility decreases if current consumption falls to less than a habitual level. The second case postulates a social dimension: there is a decline in utility if his consumption moves negatively relative to what the peers consume. This fall in utility is reasoned to be due to the perceived failure to 'keep up with the Joneses', as the old American idiom suggests. Conversely, in all cases, raising consumption above a habitual or 'Jones' level is accompanied by an increase in utility. How may this translate into a higher equity premium? One reason for the low premium in the original representative agent model is the low volatility in consumption. Such low volatility coupled with high returns on risky assets implies a very high risk aversion for agents. If, however, utility is not derived from consumption, but rather from ratios in consumption (relating current consumption to one of the above-mentioned reference levels), then even small variations in consumption may translate into more volatile, habit-adjusted levels. While there are a number of results in the literature that seem to successfully rationalize the data,[5] issues nevertheless remain. For one, in the context of habit-formation, it is possible for agents to experience welfare gains by a one-time (or periodic) lowering of the consumption level as subsequent gains ('consumption bunching') relative to that new lower habitual level would lead to increasing utility and thus quickly make

[5] Abel (1990), Constantinides (1990), Campbell and Cochrane (1999).

up for the utility lost in the period of the fall (Ljungqvist and Uhlig (1999)). In addition, these preferences can seem somewhat ad-hoc as they have no axiomatic foundations.

Market-based Solution

A promising feature of market incompleteness focuses on borrowing constraints in connection with viewing asset pricing through the lenses of a life-cycle model, as proposed by Constantinides, Donaldson, and Mehra (2002). To obtain new insights into the ongoing debate of the equity premium puzzle, they employ an overlapping generation model. New generations are born each period, which guarantees an infinite time horizon for the economy as a whole. In the first period, agents accumulate human capital while receiving low endowment income. The second period sees agents receiving a high, though stochastic wage income, whereas towards the end of their lives, in the third period, they retire and live off the wealth accumulated during the second period. The incomplete market features introduced by this model are such that no one can trade with yet unborn generations, nor can the young generation borrow against their future income, as they have no means of providing adequate collateral to secure their credit. The novelty of the approach arises from the implications derived by looking at the optimal portfolio held by each generation: The young face a situation where future wage and equity income are only weakly correlated (This is a key assumption of the model. Empirical support to the assertion is provided by Davis and Willen (2000)); investing in equity would thus be a good choice. However, not only is their income low, but with consumption smoothing on their mind, they are also extremely hesitant to spend their period-one income on anything else than consumption goods. Through the borrowing constraint, non-participation in the securities markets by the agent of the young generation arises endogenously. The middle-aged, on the contrary, have their wage uncertainties resolved, but are looking to save for old age and retirement. When they are old, they will have to live off their assets. Then, their consumption will be strongly correlated to the return of their assets. Hence they prefer to invest in assets with low variance and will buy mostly risk-free bonds. Given that the middle-aged fraction of the population provides the marginal investor, this setup provides a plausible justification for both a low return on risk-free assets as well as a very high return on equity and thus, a high equity premium. While there are still some issues such as relaxing the no-bequest assumption to be satisfyingly resolved, the idea of applying a life-cycle argument to consumption-based asset pricing proved innovative. Its outstanding contribution does not necessarily lie in the fit of its parameters but instead in its identification of a more fundamental driver of the asset-market. It can be argued that Constantinides, Donaldson, and Mehra (2002) responded to Kocherlakota's call (Kocherlakota (1996), p. 87) to transcend the "current mode of patching the standard models of asset exchange with transaction costs here and risk aversion there"−quite successfully so.

Conclusion

The importance of the equity premium puzzle is that it manifests the empirical failure of the class of general equilibrium asset-pricing models. This has been of tremendous concern ever since because these and related models are still considered a cornerstone of modern macroeconomics. Changing preference-structures can actually deepen the puzzle while at the same time highlighting the role of abstractions that contribute to it. Other attempts at modifying the utility function by incorporating habit-style preferences are successful, but their ad-hoc nature lacking axiomatic foundation needs to be resolved to gain full acceptance. A particular, encouraging proposal abandons the complete-markets representative-agent approach by introducing borrowing constraints and heterogeneity in the context of an overlapping generations model. The special appeal lies with its very intuitive yet far-reaching application of the life-cycle argument to asset pricing, identifying a fundamental feature of the market for assets. Altogether, as concluded by the discoverers of the puzzle Mehra and Prescott themselves (Mehra and Prescott (2008b), p. 114), "considerable progress has been made and the equity premium is a lesser puzzle than it was twenty years ago."

References

Abel, A. B. (1990): "Asset prices under habit formation and catching up with the Joneses," *American Economic Review, Papers and Proceedings*, 80, 38–42.

Aiyagar, S. R., and M. Gertler (1991): "Asset Returns with Transaction Costs and Uninsured Individual Risk," *Journal of Monetary Economics*, 27(3), 311–331.

Brealy, R. A., S. C. Myers, and A. Franklin (2006): *Principles of Corporate Finance*. McGraw-Hill/Irwin, New York, 8 edn.

Breeden, D. T. (1991): "An intertemporal asset pricing model with stochastic consumption and investment opportunities," *Journal of Financial Economics*, 7, 265–296.

Campbell, J. Y., and J. H. Cochrane (1999): "By force of habit: A consumption-based explanation of aggregate stock market behaviour," *Journal of Political Economy*, 107, 205–251.

Constantinides, G. M. (1990): "Habit formation: A resolution of the equity premium puzzle," *Journal of Political Economy*, 98, 519–543.

Constantinides, G. M., J. B. Donaldson, and R. Mehra (2002): "Junior can't borrow: A new perspective on the equity premium puzzle," *Quarterly Journal of Economics*, 118, 269–296.

——— (2005): "Junior is rich: bequests as consumption," *Economic Theory*, 32(1), 125–155.

Davis, S., and P. Willen (2000): "Using financial assets to hedge labor income risks: Estimating the benefits," Working Paper.

Donaldson, J., and R. Mehra (2008): "Risk-based Explanations of the Equity Premium," in *The Handbook of the Equity Risk Premium*, pp. 37–99. Elsevier, Amsterdam.

Epstein, L. G., and S. E. Zin (1989): "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis," *Journal of Political Economy*, 99, 263 – 286.

FRED (2011): "3-Month Treasury Bill: Secondary Market Rate [TB3MS]," Federal Reserve Economic Data, Federal Reserve Bank of St. Louis, Board of Governors of the Federal Reserve System.

Friend, I., and M. E. Blume (1975): "The demand for risky assets," *American Economic Review*, 65, 900–922.

Grossman, S. J., and R. J. Shiller (1981): "The determinants of the variability of stock market prices," *American Economic Review*, 71, 222–227.

Hansen, P. H., and R. Jagannathan (1991): "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99(2), 225–262.

Heaton, J., and D. Lucas (1996): "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing," *Journal of Political Economy*, 104(3), 443–487.

Homer, S. (1963): *A history of interest rates*. Rutgers University Press, New Brunswick, NJ.

Ibbotson, R. G., and R. A. Singuefield (1979): *Stocks, bonds, bills and inflation: Historical returns (1926-1978)*. Financial Analysts Research Foundation.

Kocherlakota, N. R. (1996): "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature*, XXXIV, 42–71.

Kreps, D., and E. Porteus (1978): "Temporal resolution of uncertainty and dynamic choice theory," *Econometrica*, 36, 185–200.

Ljungqvist, L., and T. J. Sargent (2004): *Recursive Macroeconomic Theory*. MIT Press, Cambridge, MI, 2 edn.

Ljungqvist, L., and H. Uhlig (1999): "On consumption bunching under Campbell-Cochrane habit formation," SFB 649 Discussion Paper, Berlin, available online at http://sfb649.wiwi.hu-berlin.de/discussionPapers de.php.

Lucas, R. E. (1978): "Asset prices in an exchange economy," *Econometrica*, 46, 1429–1445.

Mehra, R., and E. C. Prescott (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

——— (2008a): "The Equity Premium: ABCs," in *The Handbook of the Equity Risk Premium*, pp. 1–36. Elsevier, Amsterdam.

——— (2008b): "Non-Risk-based Explanations of the Equity Premium," in *The Handbook of the Equity Risk Premium*, pp. 102–114. Elsevier, Amsterdam.

Shiller, R. J. (1989, 2011): "S&P Composite Price Index Series, Dividends Accruing to Index Series, CPI Series, Market Volatility," Chapter 26, The MIT Press, available online at `http://www.econ.yale.edu/~shiller/data.htm`.

Storesletten, K., C. I. Telmer, and A. Yaron (2004): "Consumption and Risk Sharing over the Life Cycle," *Journal of Monetary Economics*, 51(3), 609–633.

Tobin, J., and W. Dolde (1971): "Wealth, Liquidity, and Consumption," Cowles Foundation Discussion Papers 311, Cowles Foundation for Research in Economics, Yale University.

Weil, P. (1989): "The equity premium puzzle and the risk-free rate puzzle," *Journal of Monetary Economics*, 24, 401–421.

Appendix

Set of admissible average risk premia and real returns

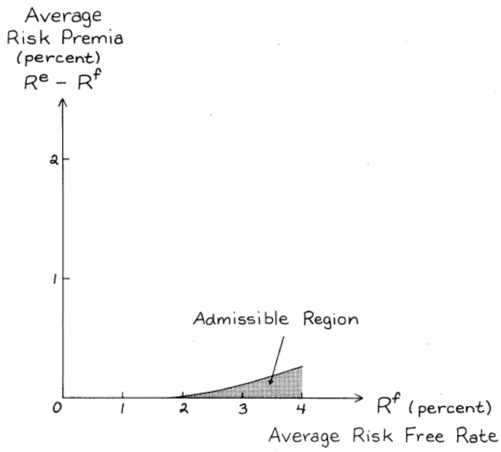

Figure 1: Set of admissible average risk premia and real returns
Source: Mehra and Prescott (1985), Fig. 4, p.155

Estimation

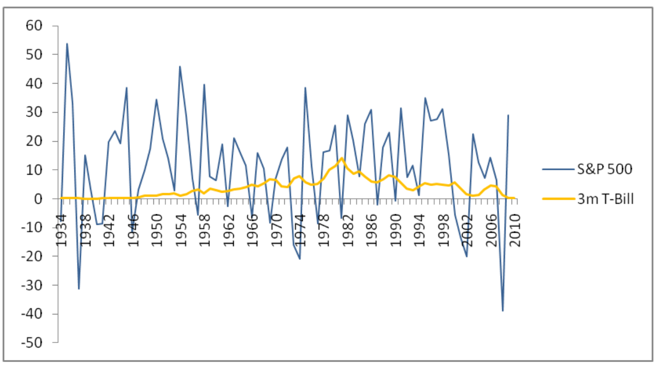

Figure 2: Mean annual total return on the S&P 500 and nominal yield on the 3-Month Treasury Bill, in percent; own calculations
Data: FRED (2011) and Shiller (1989, 2011)

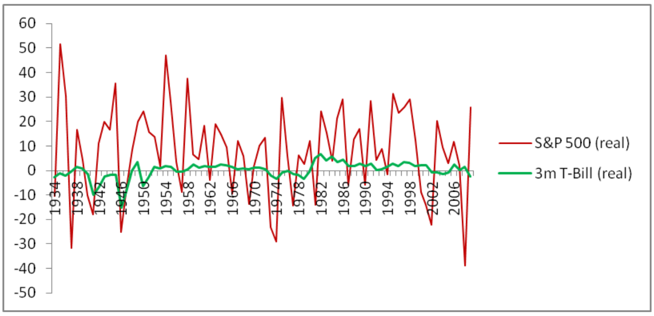

Figure 3: Mean annual real total return on the S&P 500 and real yield on the 3-Month Treasury Bill, in percent; own calculations
Data: FRED (2011) and Shiller (1989, 2011)

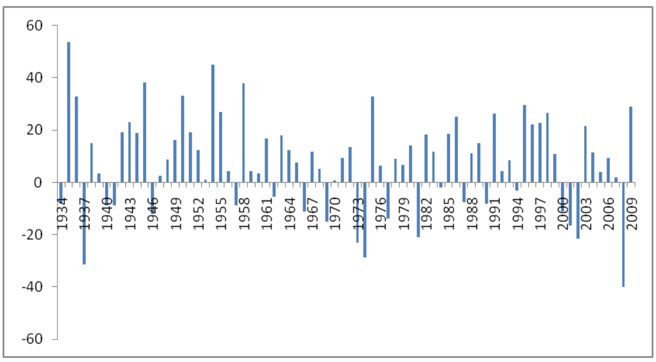

Figure 4: Annual equity premium, 1934−2010; own calculations
Data: FRED (2011) and Shiller (1989, 2011)

Essays on Herd Behavior — Theory and Criticisms

Annika Westphäling*

"Four eyes see more than two"−that information gets more precise being aggregated from more people is proverbial. Problems occur only when the aggregation of information does not run smoothly, for example if information is not communicated directly but revealed through people's actions. This is what the branch of literature studying imitative and herd-like behavior is concerned with. The concepts of herd behavior and informational cascades were first formalized in 1992 by Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992) (hereafter BHW). Herd behavior in financial markets and the resulting price bubbles had been observed since long before, without any available theoretical explanation; so when the paper by BHW was published, the model was embraced to serve as an explanation of these phenomena. However, while the results prove to be highly applicable to examples from cultural change to the spread of technology, the applicability to the functioning of financial markets is more delicate, as we will show in this essay.

The basic model involves agents acting one after another, maximizing their expected gain, the payoff depending on their decision and the underlying state. Their information consists of a private signal about the state and the actions of their predecessors. We call herd behavior the situation where a history of actions

* Annika Westphäling received her degree in Economics (B. Sc.) from the University of Bonn in 2010. The present article refers to her bachelor thesis submitted in August 2010.

leads an agent to take the same action as her predecessor, regardless of her private signal. Agents are ex-ante equal, in the sense that the precision of the signal received is the same for everybody; so when an agent starts herding, learning stops altogether, as her actions do not provide any additional information, leaving her successor in the same position where the evidence of the history outweighs any private signal. The conformal behavior of all subsequent agents is called an informational cascade.

A Classic Model of Herd Behavior

In the paper by BHW, individuals are asked whether to adopt or reject a certain behavior. Both the gain of adopting the behavior and the signal the agents receive can take either value 1 or 0, and for every agent the probability of receiving a signal that equals the real value is $p > 1/2$. The cost of adoption is constant at $1/2$ for all individuals, so an agent will adopt whenever she estimates the value being 1 more likely than the value being 0. In case of indifference, a coin flip will decide on the action. Consider the case where the first agent receives a signal 1 and adopts, thus revealing her signal. If the second agent receives a signal of 1, she will adopt; receiving a signal of 0, she is indifferent, and she adopts or rejects with equal probability.

The history (adopt; adopt) does not allow to perfectly deduce the history of signals, but indicates a high value of the behavior; indeed, it can be shown that the third agent adopts regardless of her signal, thus starting an informational cascade. If the third agent observes a history of actions (adopt; reject), she will know that the history of signals must have been $(1; 0)$, which is equally probable given a high or low value, so she finds herself in the very same position as the first agent. It follows that after an even number of agents we find ourselves in a cascade if the history features two more actions of one type than the other;

moreover, the probability of this occurring is one in the limit. The more precise the private signal, the more likely it is for a cascade on the correct action to start early on, as a precise signal favors subsequent agents receiving the same signal and thus taking the same optimal action, starting what we call a “fully revealing” cascade. However, we may not let this term mislead us: agents can only state with a certain probability whether the action they take is optimal, but learning stops as soon as the first agent engages in herd behavior.

Considering the high loss of information induced by both the coin flip and the short sequence of actions sufficing for the start of an informational cascade, it is not surprising that even with a precise signal $p >> 1/2$, the probability for a non-fully revealing cascade is very high.

Compared to the calamitous picture that is painted above, stock markets seem downright docile. In fact, both the strength and the weakness of this model lies in its simplicity: we will see that as we adapt the model to better account for the complexities of financial markets, the possibility for an inefficient outcome will be overruled.

First Critique: Modifying the action space

The first criticism to the direct application of BHW’s model to describe how financial markets operate was brought in by Lee (1993) and it concerns the richness of the action space available to our decision makers. Compared to the coarseness of the adoption or rejection of a behavior, in financial markets, it is more sensible to assume a continuous action space, as the possible quantity traded can be approximated to a continuum.

Lee allows for a finite number of states, each of which is characterized by the probability of receiving a signal 1 it induces. The action space is a subset of the interval $[0, 1]$ and it includes these probabilities. The agents have to assess the odds of receiving a signal 1 considering their signal and the actions of their predecessors; formally, they minimize the expected squared difference between their action and the underlying probability, which is their loss function.

This leads us also to alterations concerning the definition of an informational cascade: here, an informational cascade occurs whenever the history of actions converges, whether in finite time or in the limit. In the first case, the conformal action is optimal only with a certain probability, so if a cascade is fully revealing it is so by chance, whereas if actions converge only in the limit, we will see that the limes is bound to be optimal given the underlying state.

Lee states that, with any discrete action space, the emergence of a cascade as per BHW is almost certain when the number of agents goes to infinity. This follows from the observation that the longer the history, the less an agent's action will deviate from her predecessor's. For the smallest gap in the action space, there exists a history length after which an individual is no longer ready to "jump over" this gap to accommodate her signal. So after any history of that length, we find ourselves in a cascade.

Furthermore, the positive probability of a cascade is equivalent to the positive probability of an inefficient outcome. In fact, any history is induced by a sequence of private signals. Every signal has a positive probability of occurring in any state. It follows that if there exists a finite history so that a given action minimizes the expected loss independently of the signal, this history has a positive probability of occurring in any state, hence also in a state where the action is not optimal.

Moreover, note that the non-existence of such a history is sufficient for an optimal outcome. If the action somebody takes is never the same under a high as under a low signal, an agent's action reveals her private information. It follows that agents can deduce the history of private signals from the history of actions, and the strong law of large numbers states that by increasing the history length we can almost certainly deduce the true state of the world. With a continuous action space and a non-degenerate prior, an agent will always adapt her action according to her signal, albeit to an arbitrarily small degree as the length of the history grows. As we see above, this fact eliminates the possibility of herding and guarantees an efficient outcome, thus setting one major critique to the applicability of "standard" herding models to financial markets.

Second Critique: Introducing a Pricing Mechanism

Another critique to the direct applicability of BHW-like models to financial markets is founded on the assumption that in a functioning financial market, all public information is reflected in the price. In Lee, the action chosen, if informative, reveals the signal received and it equals the public belief. In the trading model by Avery and Zemsky (1998) (hereafter AZ), the estimation of the predecessor, which equals the public information, is reflected in the price, and an agent's buy or sell order reveals whether her assessment is higher or lower than her predecessor's.

Consider a financial market where agents sequentially trade one asset that can take value 1 or 0. Differently from BHW, where the cost of adoption is constant, here the price at which to buy or sell an asset depends on the history of trading. In each period, one agent from a continuum of traders is randomly selected, where λ is the fraction of informed traders, and $(1 - \lambda)$ the fraction of noise traders.

The agents may either buy or sell one unit of the stock or refrain from trading, and we assume that noise traders choose each action with equal probability; on the other hand, informed traders are risk neutral and maximize their expected profit updating their beliefs about the value of the asset after observing the history and a signal having precision p.

We need to modify the definition of herding to account for flexible prices. For

AZ, an agent engages in herd buying, if two conditions are met: first, the expected value of the asset given the initial prior has to be higher than the prior updated with her signal, so that if the agent had been the first to trade, she would have sold. Secondly, the market maker estimates the value of the asset higher than in the beginning, so the trading history must have been positive, what implies a preponderance of buying orders.

So for herding as per AZ, a signal which would have had a negative effect on the expected value of an asset in the beginning has to lift the expected value of the asset even above an ask price, which is at least as high as the estimated value of the asset given the history.

At any point in time, having observed the same trading history unfolding and having exactly the same information about it, both the traders and the market maker have the same valuation of the asset. However, once called to trade, informed traders receive a signal which moves their valuation either above or below the market maker's. The latter fixes bid and ask prices conditionally on receiving a sell and a buy order respectively. As she makes zero profits in equilibrium, and due to the presence of noise traders in the market, the market maker will always fix an ask price lower than the valuation of a trader with a high signal and a bid price higher than the valuation of a trader with a low signal. Therefore when prices are not fixed but change with the trading history and where there is only uncertainty about the value of the asset traders always follow their signal and herding is not possible, and even though the existence of noise traders blurs the information, in the long run the history of trades will reveal the true value of the asset.

To allow for herd behavior in financial markets, however, AZ offer an extension of their model, accounting for the fact that traders may be informed of events that have an impact on an asset's value, whereas the market maker is not. Traders in turn do not know for sure whether the reported event, hereafter referred to as information event, decreases or increases the value of the asset, but have to rely on the history of trades and on their private signals that work by the same principle as already established in BHW. The true value of the asset can now be either 0, $1/2$ or 1, and informed traders receive a signal $1/2$ if and only if the true value is $1/2$; otherwise, the signal is correct with probability $p > 1/2$. The market maker can never exclude $1/2$ as a possible state but at this point traders can and they interpret the trading history differently from the market maker. If the probability of an information event is small, the valuation of the market maker stays close to $1/2$ while traders' valuation can diverge to a point in which one of them could buy with a signal 0 or sell with a signal 1. Price rigidity is recreated in proportion to how close the prior probability of state $1/2$ is to 1, allowing for an arbitrarily long sequence of herd behavior that follows. During this time the traders' valuation does not move, and the market maker's valuation re-aligns with the traders' until normal trade recovers.

Note, that in this respect, herd behavior is even informationally efficient. Creating conform behavior of informed agents helps the market maker realize that the value of the asset has deviated from $1/2$. In the long run, however, the price smoothly converges to the true value unless a new shock arises. It follows that uncertainty about an information event fails to link herd behavior to catastrophic market events and extreme price distortions.

References

Avery, C., and P. Zemsky (1998): "Multidimensional Uncertainty and Herd Behavior in Financial Markets," *The American Economic Review*, 88(4), 724–748.

BANERJEE, A. (1992): "A simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 107(3), 797–817.

BIKHCHANDANI, S., D. HIRSHLEIFER, AND I. WELCH (1992): "A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100(5), 992–1027.

LEE, I. H. (1993): "On the Convergence of Informational Cascades," *Journal of Economic Theory*, 61(2), 395–411.

An Introduction to Barrier Options — Closed Form Solution and a Monte Carlo Approach

Qi-Min Fei*

Introduction

In recent years barrier options have become increasingly popular and frequently traded financial instruments, especially appearing in retail-products, so-called "certificates", broadly offered in the German retail market. Barrier options are path-dependent exotic options that become activated or null if the underlying reaches certain levels. There are four main types of barrier options that can either have call or put feature: Down-and-In, Down-and-Out, Up-and-In and Up-and-Out. The "down" and "up" refer to the position of the barrier relative to the initial underlying price. The "in" and "out" specify the type of the barrier, referring to activating and nullifying when the barrier is breached respectively. Barrier options always come at a cheaper price than ordinary options with same features (Taleb and Proß-Gill (1997)). A Down-and-Out call option for instance becomes nullified if the price of the underlying falls below the barrier. Despite being frequently traded nowadays, barrier options are still known as exotic options since they cannot be replicated by a finite combination of standard products,

* Qi-Min Fei received his degree in Economics (B. Sc.) from the University of Bonn in 2011. The present article refers to his bachelor thesis submitted in Juli 2011.

i.e. vanilla call and put options, future contracts etc. (Hausmann, Diener, and Käsler (2002)). Already in 1973, Robert C. Merton described in his article (Merton (Spring, 1973)) a closed form solution for the price of a Down-and-Out call option. Since then the market for barrier options literally exploded. This paper gives an introduction to barrier options and its properties and derives the analytic closed form solution by risk-neutral valuation. Furthermore we apply Monte Carlo simulation to derive numerical results. The great advantage of Monte Carlo simulation lies in the fact that it is robust and can be easily extended to options depending on multiple assets when no analytical solutions exist (Moon (2008)). Simple Monte Carlo simulation faces the problem that it yields both high statistical and discretization errors due to the knockout feature of the barrier option. Thus we subsequently introduce two error reduction techniques, namely *Control Variates* and *Brownian bridges* to counter these problems. Throughout the paper we assume a filtered probability space $(\Omega, \mathcal{M}, \{\mathcal{F}_t\}_{t>0}, P)$ with respect to the filtration $\{\mathcal{F}_t\}_{t>0}$ where $W = W_t$ is a standard Brownian motion. Furthermore we assume a world satisfying the Black Scholes conditions where the money market account is described by $dB_t = rB_t dt$ and the underlying S follows a geometric Brownian motion model, i.e. $dS_t = \alpha S_t dt + \sigma S_t dW_t^P$ with W_t^P denoting a standard Brownian motion under the measure P and σ and α fixed. Our final goal guiding and motivating us through the whole paper is to price a very popular German retail product: The European bonus certificate. The payoff of such a certificate with strike price K and lower boundary H is:

$$\text{Payoff} = \begin{cases} S_T, & \text{if } \exists\, t: \; S_t \leq H \text{ or } S_T > K \\ K, & \text{else} \end{cases}$$

The payoff of such an ordinary bonus certificate is shown in Figure 1. Investors see bonus certificates as alternatives to direct investments into the underlying.

They offer the holder the chance to earn more than holding the underlying as long as the underlying stays between a strike price K and a boundary H with $K \geq H$. Those products are primarily purchased by investors who believe that the underlying will not fluctuate a lot. If this expectation turns out to be true the bonus certificate will yield a higher payoff than the underlying. A bonus certificate is a portfolio consisting of a zero-strike call and a long position in a Down-and-Out put option (Reinmuth (2002)): $p_{bc} = p_{dkop} + p_{zero\ call}$ where p_{bc} is the price of a bonus certificate, p_{dkop} that of a Down-and-Out put option and $p_{zero\ call}$ that of a zero-strike call option. Thus we regard a simple European Down-and-Out put option in the following. Nonetheless we keep in mind that there also exist "non-simple" barrier options such as multi-barrier options or barrier options that require the asset price to not only cross a barrier, but spend a certain length of time across the barrier in order to knock in or knock out. The analytical challenge in our case is to calculate p_{dkop}. Therefore we apply the technique of risk-neutral valuation and make use of deep results from stochastic calculus such as the *Reflection Principle* and *Girsanov's theorem* to calculate the expectation of the payoff under the risk-neutral measure and discount it with the risk-free spot rate similar to pricing a vanilla put option (Steele (2001)):

$$p_{dkop} = e^{-rT} E^Q \left[(K - S_T) \mathbf{1}_{\{K \geq S_T; \inf_{t \in [0;T]} S_t \geq H\}} \right]$$

with time to maturity T, strike K, barrier $H \leq K$ and Q a risk-neutral measure with the money market account as numéraire. In our numerical simulation part we simulate our asset price according to geometric Brownian motion and implement the barrier as nullifying condition for each path. Furthermore we introduce a variance reduction technique called Control Variates and a discretization error reducing technique exploiting the idea of Brownian bridges. We conclude with a discussion on the scope and limitations of the introduced techniques. The specific bonus certificate that we price in this paper is a Goldman Sachs certificate with ISIN DE000GS3DWL0 Goldman Sachs (2011a) and properties shown in Table 1 on May 8, 2011. The ask price of this European bonus certificate is **84.06** at day of pricing, so this price acts as our benchmark after taking into account limits of our model like the issuer risk and profit margins of Goldman Sachs.

Analytical Solution

Given linearity of the expectation the price of our option can be written as:

$$\begin{aligned} p_{dkop} &= e^{-rT} E^Q \left[(K - S_T) \mathbf{1}_{\{K \geq S_T ; \inf_{t\in[0;T]} S_t \geq H\}} \right] \\ &= e^{-rT} \Big(E^Q \left[(K - S_T) \mathbf{1}_{\{S_T < K\}} \right] - E^Q \left[(K - S_T) \mathbf{1}_{\{S_T < H\}} \right] \\ &\quad - E^Q \left[(K - S_T) \mathbf{1}_{\{S_T > H ; \inf_{t\in[0;T]} S_t < H\}} \right] \\ &\quad + E^Q \left[(K - S_T) \mathbf{1}_{\{S_T > K ; \inf_{t\in[0;T]} S_t < H\}} \right] \Big) \end{aligned}$$

We immediately see that the first expectation is just the price of a plain-vanilla European put option with strike K that we know from Black (1973), yielding $e^{-rT} E^Q \left[(K - S_T) \mathbf{1}_{\{S_T < K\}} \right] = K e^{-rT} N(-d_2) - S_0 N(-d_1)$, with $d_1 = \sigma\sqrt{T} + \frac{ln(\frac{S_0}{K}) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$ and $d_2 = d_1 - \sigma\sqrt{T}$. Applying the same risk-neutral valuation technique to the second expectation it takes us little effort to see that we only need to replace K by H in d_1 and d_2 to get the second expectation. Now we turn to the third and fourth term. Exemplarily we calculate the fourth expectation setting $m := \frac{r - \frac{1}{2}\sigma^2}{\sigma}$, $h := \frac{1}{\sigma} \ln\left(\frac{H}{S_0}\right)$ and $k := \frac{1}{\sigma} \ln\left(\frac{K}{S_0}\right)$ for sake of readability. We plug in $S_t = S_0 \exp\{(r - \frac{1}{2}\sigma^2)t + \sigma W_t^Q\}$ with $W_t^Q = W_t^P + \frac{\alpha - r}{\sigma} t$ being a Brownian motion under the measure Q (Hull (2007)) for S_T and S_t to get:

$$
\begin{aligned}
&e^{-rT}E^Q\left[(K-S_T)\mathbf{1}_{\{S_T>K;\inf_{t\in[0;T]}S_t<H\}}\right]\\
&=e^{-rT}\Bigg(KE^Q\left[\exp\left\{mW_T^Q-\frac{1}{2}m^2T\right\}\mathbf{1}_{\{W_T^Q>k;\inf_{t\in[0;T]}W_t^Q<h\}}\right]\\
&-E^Q\left[\exp\left\{mW_T^Q-\frac{1}{2}m^2T\right\}S_0\exp\left\{\sigma W_T^Q\right\}\mathbf{1}_{\{W_T^Q>k;\inf_{t\in[0;T]}W_t^Q<h\}}\right]\Bigg)
\end{aligned}
$$

Now we calculate the joint distribution $Q(W_T^Q>k;\inf_{t\in[0,T]}W_t^Q<h)$ dismissing the *inf term* making use of the reflection principle. Plugging back m,h and k we get the desired expression:

$$
\begin{aligned}
&=e^{-rT}K\left(\frac{H}{S_0}\right)^{\frac{2r-\sigma^2}{\sigma^2}}N\left(\frac{\ln\left(\frac{H^2}{S_0K}\right)+rT\frac{1}{2}\sigma^2T}{\sigma\sqrt{T}}\right)\\
&-S_0\left(\frac{H}{S_0}\right)^{\frac{2r+\sigma^2}{\sigma^2}}N\left(\frac{\ln\left(\frac{H^2}{S_0K}\right)+rT+\frac{1}{2}\sigma^2T}{\sigma^2\sqrt{T}}\right)
\end{aligned}
$$

Finally we can put all four expectations together and get the price:

$$
\begin{aligned}
\mathbf{p_{dkop}}&=Ke^{-rT}N(-d_2)-S_0N(-d_1)+S_0N(-x_1)\\
&-Ke^{-rT}N(-x_1+\sigma\sqrt{T})-S_0\left(\frac{H}{S_0}\right)^{2\lambda}[N(y)-N(y_1)]\\
&+Ke^{-rT}\left(\frac{H}{S_0}\right)^{2\lambda-2}[N(y-\sigma\sqrt{T})-N(y_1-\sigma\sqrt{T})],
\end{aligned}
\tag{1}
$$

with $\lambda:=\frac{r+\frac{\sigma^2}{2}}{\sigma^2}$, $y:=\frac{\ln\left(\frac{H^2}{SK}\right)}{\sigma\sqrt{T}}+\lambda\sigma\sqrt{T}$, $x_1:=\frac{\ln\left(\frac{S}{H}\right)}{\sigma\sqrt{T}}+\lambda\sigma\sqrt{T}$, $y_1:=\frac{\ln\left(\frac{H}{S}\right)}{\sigma\sqrt{T}}+\lambda\sigma\sqrt{T}$. Now let us turn back to our bonus certificate. By letting K going to zero we calculate the price of our zero-call on the DAX using the Black-Scholes formula for European call options (Hull (2007)) as 74.9225. Plugging the data of our bonus certificate into the above derived formula (1) for pricing European Down-and-Out put options we get: $\mathbf{p_{dkop}}=9.4625$. In summing up the two prices we get the final price of the European bonus certificate as **84.39** which is close to the quoted price of **84.06**.

Numerical Simulation

For our Monte Carlo simulation we generate M underlying asset price paths $\{(S^1_{t_1}, ..., S^1_{t_n}), ..., (S^M_{t_1}, ..., S^M_{t_n})\}$ on a fixed set of points in time $0 < t_1 < ... < t_n = T$ for $k = 1, ..., M$ each following a geometric Brownian motion. We then determine the payoff of the barrier option conditional on the fact that no barrier-breach has occurred at any time step. Subsequently we take the discounted average of the payoffs to obtain the numerical price of our barrier option. This estimator is unbiased and converges with probability 1 as $n \to \infty$ and the statistical error is of order $\mathcal{O}(1/\sqrt{M})$. For Barrier Options with continuous knock-out observation the discretization error−here the hitting time error (i.e. missing a barrier-breach that happens between two simulated time steps)−is of order $\mathcal{O}(1/\sqrt{n})$ (Gobet (2000)). The algorithm is presented in Algorithm 1. In our case we choose a discretization of N = 1,000 equidistant time steps per year and simulate M = 20,000 paths in total for our Monte Carlo simulation. The resulting price is **9.4976** compared to the theoretical price of **9.4625**. The difference in value can firstly be explained by the fact that we have a very high variance of **89.06** (9.44 standard deviation) due to the knockout feature of the option. Secondly we have a discretization of only 1,000 time steps, so the option can only knockout on those nodes whereas our initial barrier option theoretically can knock out at any time. In the following we will tackle the first problem by reducing the variance of our results, the second problem will be addressed thereafter through the concept of Brownian bridges, where we estimate the probability of the option knocking out between two time steps.

Control Variates

After having introduced the simple Monte Carlo simulation we now turn our focus to improving the efficiency of our simulation. In the following we introduce a variance reducing technique called Control Variates (Glasserman (2003)). The idea behind this technique is to exploit information about the errors in estimates of known variables that have a dependence to our variable (in our case $\text{Payoff}^k := Y^k$) and thus reduce the variance of our variable. We therefore choose another output X_k (*the control variate*) that is correlated to our payoff and of which we know the expectation and regard $Y^k(b) = Y^k - b(X_k - E[X])$ for a fixed b. By minimizing $Var[Y^k(b)]$ we find that the optimal choice of b^* is given by $b^* = \frac{\sigma_Y}{\sigma_X}\rho_{XY} = \frac{Cov[X,Y]}{Var[X]}$, where ρ_{XY} stands for the correlation between X and Y. Subsequently we estimate $Cov[X,Y]$ and $Var[X]$ using least-squares regression to estimate b^* and arrive at our control variate estimator $\overline{Y}(b) = \overline{Y} - \hat{b}_M(\overline{X} - E[X]) = \frac{1}{M}\sum_{k=1}^{M}(Y^k - \hat{b}_M(X_k - E[X]))$. This estimator is consistent and unbiased if the correct b^* is known. The estimation of b^* however introduces some bias that vanishes quickly as M becomes sufficiently large (Glasserman (2003)). The algorithm is described in Algorithm 2. We start with a primitive control setting X_k as *iid* normally distributed random variables, then move forward to use the underlying asset as control variate and finally we examine a plain vanilla option as control variate. As our crude Monte Carlo paths are simulated from a sequence of independent standard normal random variables there exists some correlation between those random variables and our desired payoff. The correlations between underlying and payoff as well as between the plain vanilla option and payoff are straightforward. Implementing the primitive control in our initial model we find out that we get an option price of **9.4976** and the variance becomes **89.0632640717** (9.44 standard deviation) compared to **89.0632748101** as before. The variance reduction is minimal as we have very small correlation. But those generic control variates are always available in a simulation and are mostly easy to implement. In contrast, using the underlying price as control variate yields a variance of **17.02** (4.13 standard deviation) which is significantly better than the primitive control. The correlation and thus the effectiveness of the control variate depends on the strike and the barrier. When we regard our Down-and-Out put option we can imagine that the higher the strike is the greater $|\rho_{SY}|$ becomes. Also the lower our barrier is the more we expect the correlation to be high speaking in absolute terms. Furthermore we observe that with plain vanilla put options as control variate our model yields a variance of **2.9934e-24** (1.73e-12 standard deviation) and the resulting price does not differ from our theoretical value up to the **fifth decimal**. This kind of control is very effective if the barrier is low. In fact, if the barrier is set to be zero the correlation is very close to 1. Simulation results are summarized in Table 3.

Brownian Bridges

Now that we have introduced a powerful technique reducing the statistical error of our simulation we turn our focus to reducing the discretization error, in our case the hitting-time error. Hitting-time error refers to the error that arises from not sufficiently fine discretization, i.e. when a continuously observed barrier option knocks out between two simulated time steps, but the underlying asset subsequently recovers at the latter time step, formally: For simulated time steps t_i and t_{i+1} we have $S_{t_i} > H$ and $S_{t_{i+1}} > H$, but $S_t < H$ for at least one $t \in [t_i, t_{i+1}]$. If we cannot simulate a large number of time steps due to limited computational time or resources the hitting-time error can become substantially large. Inspired by Mannella (1999) Moon (2008) proposed in his paper an efficient technique using Brownian bridges and the uniform distribution to reduce this hitting-time error for an Up-and-Out call option. The idea is to calculate an exit probability for each pair of time steps, i.e. the probability that the underlying breaches a lower boundary between two time steps. We then use a uniformly distributed random variable to decide whether this probability is sufficiently large to let our option knock out. Formally we regard a domain $D = (H, \infty)$ and define the probability P_i that the process S exits D at $t \in [t_i, t_{i+1}]$ given that S_{t_i} and $S_{t_{i+1}}$ are in D. Then P_i is exactly the exit probability that we have described above. To calculate this exit probability we can use the law of Brownian bridges and get:

$$P_i = P\left[\min_{t\in[t_i,t_{i+1}]} S_t \leq H \middle| S_{t_i} = s_1, S_{t_{i+1}} = s_2\right] = \exp\left(-2\frac{(H-s_1)(H-s_2)}{\sigma^2(t_{i+1}-t_i)}\right)$$

with s_1 and s_2 in D. In our algorithm we sample for each time step t_i a uniformly distributed random variable $u_i \sim \mathcal{U}(0,1)$. If in t_i the exit probability P_i exceeds u_i we dismiss the path and set the payoff as zero. The algorithm with Brownian bridges is presented in Algorithm 3. As the barrier option that we have discussed until now has a very low barrier, thus our simulation with control variates yields a very good result, we want to illustrate the Brownian bridge approach in an example where the barrier is tight. We regard a Down-and-Out barrier put option on the DAX with the properties in Table 2 again priced on May 8, 2011. Using our analytic formula (1) we get a theoretical price of **0.4313**. In this case the barrier is only about 7% below the actual price. As we can imagine the simulated price using simple Monte Carlo should be quite bad since it is very likely that the option undetectedly knocks out between two time steps. Again simulating with 20,000 paths and 1,000 time steps we see a price of **0.4936** for our barrier option which is more than 14% above the theoretical price. In contrast, applying the Brownian bridge technique introduced before we get a price of **0.4484** which represents a significant improvement against the simple

Monte Carlo simulation. Again summarized results are presented in Table 3.

Concluding Remarks

Using the example of a European bonus certificate we examined in this paper basic properties and the pricing of barrier options both analytically and numerically. We practically introduced two techniques to reduce variance of the simple Monte Carlo simulation on the one hand and to reduce discretization error on the other hand. We observe that our analytical price is different from the quoted price in the market. Several factors lead to the possible discrepancy. Adjusting for issuer risk, volatility skew, barrier shift and risk-free rate we expect to gain more accurate results. Also given our foundation it is not a difficult task to extend our model to pricing more complex barrier options, such as multi-barrier, Asian or even Parisian barrier options. For loose barriers we propose reducing variance by using appropriate control variates and going further various other numerical techniques like for instance importance sampling. For relatively tight barriers we propose reducing discretization error by using the Brownian bridge approach.

References

Baldi, P. (1995): "Exact asymptotics for the probability of exit from a domain and applications to simulation," *The Annals of Probability*, 23(4), 1644–1670.

Bjoerk, T. (2004): *Arbitrage Theory in Continuous Time, Second Edition.* Oxford University Press, Oxford.

Black, F. (1973): "The Pricing of Options and Corporate Liabilities," *The Journal of Political Economy*, 81(3), 637–654.

Bloomberg (2011): "1 Year German Government Bond," Available online at `http://www.bloomberg.com/markets/rates-bonds/government-bonds/germany/`. Accessed May 8, 2011.

De Weert, F. (2008): *Exotic Options Trading*. John Wiley and Sons, Chichester.

Glasserman, P. (2003): *Monte Carlo Methods in Financial Engineering*. Springer, New York.

Gobet, E. (2000): "Weak approximation of killed diffusion using Euler schemes," *Stochastic Process*, 87(2), 167–197.

Goldman Sachs (2011a): "Bonus Certificate DE000GS3DWL0," Available online at `http://www.goldman-sachs.de/isin/detail/DE000GS3DWL`. Accessed May 8, 2011.

——— (2011b): "Implied Volatility of DAX," Available online at `http://www.goldman-sachs.de/isin/detail/DE000GS3Z2X9/`. Accessed May 8, 2011.

——— (2011c): "Payoff of a bonus certificate," Available online at `http://www.goldman-sachs.de/isin/detail/DE000GS3DWL0/`. Accessed May 8, 2011.

——— (2011d): "Rating and CDS Spread," Available online at `www.gs.de/default/default/rating/navid,276/lq_switch,0/`. Accessed May 8, 2011.

Hausmann, W., K. Diener, and J. Käsler (2002): *Derivate, Arbitrage und Portfolio-Selection. Stochastische Finanzmarktmodelle und ihre Anwendungen.* Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig, Wiesbaden.

Hull, J. (2007): *Options, Futures and Other Derivatives 6th Edition*. Prentice-Hall of India, New Delhi.

Igusa, K. (1990): "Brandeis University," Available online at `http://people.brandeis.edu/~igusa/Math56aS08/Math56a_S08_notes082.pdf`. Accessed May 8, 2011.

Jeanblanc, M., M. Yor, and M. Chesney (2009): *Mathematical Methods for Financial Markets*. Springer, Berlin.

Mannella, R. (1999): "Absorbing boundaries and optimal stopping in a stochastic differential equations using exponential timestepping," *Physics Letters*, pp. 257–262.

Merton, R. C. (Spring, 1973): "Theory of Rational Option Pricing," *The Bell Journal of Economics and Management Science*, 4(1), 141–183.

Moon, K. (2008): "Efficient Monte Carlo Algorithm for Pricing Barrier Options," *Communications of the Korean Mathematical Society*, 23(2), 285–294.

Reinmuth, P. (2002): *Barrier Options: Merkmale, Bewertung, Risikomanagement und Anwendungsmöglichkeiten.* Haupt Verlag, Bern, Stuttgart, Wien.

Steele, J. (2001): *Stochastic Calculus and Financial Applications*. Springer-Verlag, New York.

Taleb, N. N., and I. Pross-Gill (1997): *Dynamic Hedging: Managing Vanilla and Exotic Options*. Wiley Finance, New York.

Appendix

Algorithms

Algorithm 1. ***Standard Monte Carlo Method***

for $k = 1, ..., M$

for $i = 1, ..., n$

generate a $\mathcal{N}(0,1)$ *sample* Z_i

set $S^k_{t_{i+1}} = S^k_{t_i} \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)\sqrt{t_{i+1} - t_i} + \sigma\sqrt{t_{i+1} - t_i}\ Z_i\right\}$

end

if $\max_{1 \leq i \leq n} S_{t_i} > H$ *then* $Payoff^k = \max(K - S^k_{t_n})$

else $Payoff^k = 0$

end

set $p_{dkop} = \frac{1}{M}\sum_{k=1}^{M}\left(e^{-rT} Payoff^k\right)$

Algorithm 2. ***Control Variate Monte Carlo Method***

for $k = 1, ..., M$

generate X_k

for $i = 1, ..., n$

generate a $\mathcal{N}(0,1)$ *sample* Z_i

set $S^k_{t_{i+1}} = S^k_{t_i} \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)\sqrt{t_{i+1} - t_i} + \sigma\sqrt{t_{i+1} - t_i}\ Z_i\right\}$

end

if $\max_{1 \leq i \leq n} S_{t_i} > H$ *then* $Y^k = \max\left(K - S^k_{t_n}\right)$

else $Y^k = 0$

end

set $\hat{b}_M = \frac{\sum_{k=1}^{M}(X_k - \overline{X})(Y^k - \overline{Y})}{\sum_{k=1}(X_k - \overline{X})^2}$

for $k = 1, ..., M$

set $Y^k = Y^k - \hat{b}_M(X_k - E[X])$

end

set $p_{dkop} = \frac{1}{M}\sum_{k=1}^{M}\left(e^{-rT}Y^k\right)$

Algorithm 3. ***Brownian Bridges***

for $k = 1, ..., M$

for $i = 1, ..., n$

generate a $\mathcal{N}(0,1)$ *sample* Z_i

set $S^k_{t_{i+1}} = S^k_{t_i} \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)\sqrt{t_{i+1} - t_i} + \sigma\sqrt{t_{i+1} - t_i}\ \ Z_i\right\}$

set $P_{t_{i+1}} = \exp\left\{-2\frac{(H - S_{t_i})(H - S_{t_{i+1}})}{\sigma^2(t_{i+1} - t_i)}\right\}$

end

generate a $\mathcal{U}(0,1)$ *sample* $u_i, i = 1, ..., n$

if $S_{t_i} > H$ *and* $P_{t_i} < u_i, \forall i \ \ 1 \leq i \leq n$ *then* $Y^k = \max\left(K - S^k_{t_n}\right)$

else $Y^k = 0$

end

set $p_{dkop} = \frac{1}{M}\sum_{k=1}^{M}\left(e^{-rT}Y^k\right)$

Figures

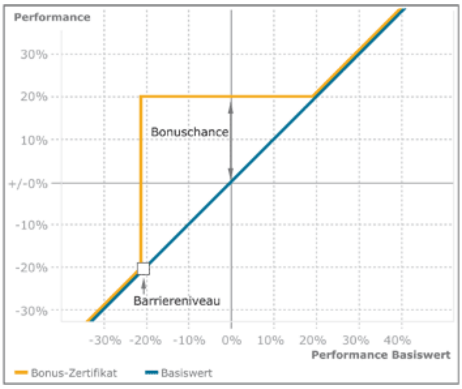

Figure 1: Payoff of a bonus certificate
Source: Goldman Sachs (2011c)

Tables

The underlying prices, the levels of the barriers and the exercise prices have been multiplied by 0.01.

Underlying	DAX Performance Index
Time to maturity	11 months
Exercise price (K)	82.5
Barrier (H)	27
Price of the underlying	74.9225
Credit Rating of the Issuer	A+/A1 (Fitch/Moody's) Goldman Sachs (2011d) CDS +71,08bp
Risk-free interest rate	1.38% (one year German government bond) Bloomberg (2011)
Implied volatility	18.2071% (annualized implied volatility of the DAX) Goldman Sachs (2011b)

Table 1: Bonus Certificate

Underlying	DAX Performance Index
Time to maturity	1 Year
Exercise price (K)	82.5
Barrier (H)	70
Price of the underlying	74.9225
Risk-free interest rate	1.38% (one year German government bond) Bloomberg (2011)
Implied volatility	18.2071% (annualized implied volatility of the DAX) Goldman Sachs (2011b)

Table 2: Second Barrier Option

First Barrier Option	**Option Price**	**Bonus Certificate**	**Variance**
Quoted	NA	84.06	-
Analytical	9.4625	84.39	-
Simple MC Simulation	9.4976	84.42	89.06
Primitive Control	9.4976	84.42	89.06
Underlying as Control	9.4906	84.41	17.02
Vanilla Option as Control	9.4625	84.39	2.99e-24
Second Barrier Option			
Analytical	0.4313	-	-
Simple MC Simulation	0.4808	-	3.02
Brownian Bridge Monte Carlo	0.4465	-	2.68

Table 3: Summary of Results

Testing Normality

Hanno Förster*

Introduction

As stated in the well known Lindeberg-Levy Central Limit Theorem (CLT) any sum of standardized *iid* random variables that have finite variance converges in distribution to the standard normal, a result that is commonly used in Econometrics to approximate distributions of test statistics and estimators. Yet, there are a number of cases where neither the Lindeberg Levy Theorem nor similar CLTs can be invoked. E.g. when large samples are unavailable, one cannot rely on asymptotics in order to approximate the distribution of the least squares estimator in the linear model. In this instance the error terms are commonly assumed to be normally distributed to make the derivation of the estimator's distribution feasible. Likewise maximum likelihood estimation requires knowledge of the exact distribution of the error term and hence frequently relies on the assumption that a random sample is independently $\mathcal{N}(0, \sigma^2)$ distributed. For these and similar cases it is desirable to have a statistical test to assess if normality is a reliable assumption. Importantly since the need for testing normality sometimes arises in the first place because large samples are unavailable (like in the linear model), finite sample performance deserves special attention when assessing the viability of such tests.

* Hanno Förster received his degree in Economics (B. Sc.) from the University of Bonn in 2011. The present article refers to his bachelor thesis submitted in September 2011.

Arguably the most cited and most commonly applied examples in the existing literature on testing normality are the Kolmogorov Smirnov test and the Jarque Bera test (JB test). The Kolmogorov Smirnov test proposed in Kolmogorov (1933) and Smirnov (1948) is a nonparametric test, which evaluates the sample cumulative distribution function. The JB test proposed in Jarque and Bera (1980) and Jarque and Bera (1987) is based on third and fourth sample moments.

The objective of my work is to draw a comparison between the ‘traditional’ JB test and an approach that has been proposed in Bontemps and Meddahi (2005) and which is based on the Generalized Method of Moments (GMM) framework. In particular I present a theoretical derivation of each test’s statistic and its distribution, a simulation based assessment of their finite sample properties and a brief conclusion on the merits and demerits of each test.

In this summary I present a complete derivation of the JB test’s distribution, a sketch of the derivation of the GMM approach, a summary of the simulation results and a brief conclusion on the viability of each test.

Derivation of the Asymptotic Distribution of the Jarque Bera Statistic

In Jarque and Bera (1980) the JB test is developed by applying the lagrange multiplier principle to the pearson family of distributions. In contrast I present in the following a derivation of the test statistic’s asymptotic distribution by straightforward application of the delta method.[1]

Let X_i denote the random Variable of interest. For an *iid* sample $X_1, ..., X_n$ the JB-statistic is defined

$$JB \quad = \quad n \left(\frac{1}{6} \, Skewness^2 + \frac{1}{24} \; (Kurtosis - 3)^2 \right),$$

[1] See Van der Vaart (2007) for details on the Delta Method.

where

$$Skewness = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^3}{s^3}, \quad Kurtosis = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^4}{s^4},$$

and $\overline{X}$, s are sample mean and sample standard deviation respectively. In order to determine the asymptotic distribution of $Skewness$ and $Kurtosis$ (and ultimately of JB) define

$$g(a,b,c,d) := \frac{c-3ab+2a^3}{(b-a^2)^{\frac{3}{2}}}, \qquad h(a,b,c,d) := \frac{d-4ac+6ba^2-3a^4}{(b-a^2)^2},$$
$$y_n := \left(\overline{X}, \overline{X^2}, \overline{X^3}, \overline{X^4}\right)^T,$$

such that $Skewness = g(y_n)$ and $Kurtosis = h(y_n)$. The Delta Method states that for any function f that is continuously differentiable at μ

$$\sqrt{n}(y_n - \mu) \xrightarrow{d} \mathcal{N}(0,\Sigma) \quad \text{implies} \quad \sqrt{n}\left(f(y_n) - f(\mu)\right) \xrightarrow{d} \mathcal{N}\left(0, f'(\mu)\Sigma f'(\mu)^T\right).$$

Note that if the sample is indeed taken from the standard normal, we have[2]

$$\sqrt{n}\begin{pmatrix} \overline{X} \\ \overline{X^2} - 1 \\ \overline{X^3} \\ \overline{X^4} - 3 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0,\Sigma^0) \quad \text{where} \quad \Sigma^0 = \begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 2 & 0 & 12 \\ 3 & 0 & 15 & 0 \\ 0 & 12 & 0 & 96 \end{pmatrix}$$

and

$$g(\mu^0) = 0\,,\; g'(\mu^0) = (-3,0,1,0)\,,\; h(\mu^0) = 3\,,\; h'(\mu^0) = (0,-6,0,1)\,.$$

[2] Population moments of the standard normal can be found in Table 1 in the Appendix.

Application of the delta method to g and h (both differentiable at $\mu^0 = (0,1,0,3)^T$) yields that for a sample drawn from the standard normal $\sqrt{n}$ *Skewness* $\xrightarrow{d} \mathcal{N}(0,6)$ and $\sqrt{n}$ *Kurtosis* $\xrightarrow{d} \mathcal{N}(0,24)$ and hence ultimately $JB \xrightarrow{d} \chi^2_{(2)}$.

The GMM Approach, a brief Outline

The derivation of the approach proposed in Bontemps and Meddahi (2005) requires introduction of some terminology. The toehold of GMM estimation is a *population moment condition* $E\left(m(X_i)\right) = 0$ where m is a vector of functions $m_1, \dots, m_l$ and X_i is an observed random variable.[3] Since the validity of this condition is a crucial assumption in GMM estimation, it is often desirable to test if the population moment condition is indeed satisfied. This can be done using a *test for overidentifying restrictions*

$$J = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(X_i)^T \, W^* \, \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(X_i) \quad \text{where} \quad W^* = \, E\left(m(X_i) m(X_i)^T\right)^{-1}.$$

If the population moment condition holds, we have $J \xrightarrow{d} \chi^2_{(l)}$. Replacing W^* by a consistent estimate leaves the asymptotic distribution of J unaffected.

Bontemps and Meddahi (2005) combine these concepts with a relationship known as the Stein equation, namely that $X \sim \mathcal{N}(0,1)$ if and only if $E(\, q'(X) - Xq(X)\,) = 0$ for all continuously differentiable q that satisfy $\int |q'(x)| e^{-x^2/2} \mathrm{d}x < \infty$.[4] With the help of the Stein equation for any suitable function q_j we can define functions $m_j(X) =: q_j'(X) - Xq(X)$ that stacked into a vector $m(X)$ satisfy a population moment condition if and only if X is standard normal. The test for overidentifying restrictions thusly turns into a test for normality. Note that interestingly by inserting Skewness and Excess Kurtosis as population moment

[3] Note m does not depend on unobserved parameters. In fact this is a special case of a moment condition that is interesting for our purpose.

[4] The Stein equation was derived in Stein (1972).

conditions, the JB test as well can be motivated as special case of the GMM approach to testing normality.

As appealing choice of functions q_j Bontemps and Meddahi (2005) suggest Hermite polynomials

$$H_j(X) = \frac{(-1)^j}{\sqrt{j!}} e^{\frac{X^2}{2}} \frac{\partial^j}{\partial X^j} e^{\frac{-X^2}{2}}.$$

These polynomials have some useful properties. For instance they are orthonormal, i.e. $\mathbf{E}\left(H_j(X)H_k(X)\right) = \mathbf{I}_{(j=k)}$ and satisfy the Stein Equation if and only if $\mathbf{E}[H_j(X)] = 0$ for all $j > 0$.[5] For one thing these properties simplify the computation of the test statistic. Inserting a selection of Hermite polynomials $\{H_j\}_{j \in I}$ yields

$$J_H = \frac{1}{n} \sum_{j \in I} \left(\sum_{i=1}^{n} H_j(X_i) \right)^2 .$$

Apart from causing computational convenience the properties of the Hermite Polynomials prove to be very beneficial in situations where the unobserved variable of interest is not directly observed. This case arises for instance in testing normality of the unobserved error term in the linear model. Although it seems natural to replace the variable of interest by a consistent estimate (e.g. regression residuals in the above mentioned example), this can distort the distribution of the test statistic, a fact that is sometimes refered to as *parameter uncertainty problem*. In order to formalize this problem consider a vector m that depends on a random vector X of observed data, but also on an unobserved parameter vector θ^0 that can be estimated consistently by $\hat{\theta}$. A first order Taylor expansion

[5] In my work I fill in the details to a proof presented in Bontemps and Meddahi (2002).

of m around θ^0 yields

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(X_i,\hat{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(X_i,\theta^0) + \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(X_i,\theta^0)}{\partial\theta^T}\right)}_{=:\mathcal{D}}\sqrt{n}\left(\hat{\theta}-\theta^0\right) + \epsilon\,.$$

It is easy to see that the solution to the problem can be boiled down to finding f such that $\lim_{n\to\infty}\mathcal{D}=0$. Bontemps and Meddahi (2005) show that in a wide range of cases of parameter uncertainty, including the linear regression example described above, the Hermite Polynomials satisfy this condition.[6] The key to this feature lies in a combination of the orthonormality property with other properties that are specific to Hermite Polynomials. Thus by inserting Hermite polynomials into the test statistic for overidentifying restrictions we get a test for normality that is robust towards parameter uncertainty.

Simulation Evidence

As emphasized earlier, the necessity for testing normality often arises in the first place, because large samples are unavailable. Thus in testing normality finite sample performance is of special importance. In my work I assess the finite sample properties of the considered approaches by the means of simulation. In particular I revisit simulations presented in Bontemps and Meddahi (2005) and consider additional settings that seemed to hold promise to reveal advantages and disadvantages of the considered tests. A selection of results is presented in Tables 2 through 7 in the Appendix.

Like Bontemps and Meddahi (2005) I find that all considered tests exhibit good size properties, although most tests based on the GMM approach tend to overreject a little bit, in general they still it perform better than the JB test that underrejects a bit, especially when applied to samples of 100 observations

[6] For a proof see Bontemps and Meddahi (2005).

or less. Both tests are very powerful against a strongly skewed exponential distribution ($\lambda = 1$). Against t-distributions in contrast all considered tests exhibit unsatisfyingly low power.

Moreover I apply both approaches to Bimodal distributions. For the considered Bimodal distributions the GMM approach is outperformed by the JB test by a bit, if even Hermite polynomials are included and by far, if used exclusively with odd Hermite Polynomials.

Note that throughout odd Hermite Polynomials seem to be good at detecting skewed distributions whereas even Hermite Polynomials seem to be best at detecting deviations in curvature.

Conclusion

All things considered, both approaches to testing normality seem to do a good though not excellent job. The JB test works well when applied to large samples, but does not exhibit entirely satisfying size properties when applied to small samples. Moreover it has poor power against symmetric alternatives like t- or bimodal distributions.

Bontemps and Meddahi (2005) contribute an insightful framework for testing normality that shows a new perspective on preceding approaches and holds promise to be a useful basis for future research in this and related subfields. The specifically proposed tests for normality exhibit good size properties, but have poor power against several alternatives. Ultimately the GMM approach can certainly not be said to perform generally better than the JB test. It might be of interest however, that if a specific alternative is suspected, a test based on carefully selected Hermite polynomials is likely to perform better than the JB test.

I conclude by remarking that although it is true that the shortcomings that both tests exhibit when applied to samples of 100 observations or less make them unreliable for testing normality in small samples, in instances where large samples are available (and yet the usual CLTs do not spare us the test), both discussed approaches make well suited candidates for testing normality.

References

Bontemps, C., and N. Meddahi (2002): "Testing Normality: a GMM Approach," *Working Paper, University of Montreal*, 10, 12–200.

——— (2005): "Testing Normality: a GMM Approach," *Journal of Econometrics*, 124, 149–186.

Casella, G., and R. L. Berger (2002): *Statistical Inference*. Duxbury.

Davidson, R., and J. G. MacKinnon (2004): *Econometric Theory and Methods*. Oxford University Press.

Eichelsbacher, P. (2003): "Die Steinsche Methode," Manuscriript. Ruhr Università Bochum.

Greene, W. H. (2008): *Econometric Analysis*. Pearson Education.

Hall, A. (2005): *Generalized Method of Moments*. Oxford University Press.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Jarque, C., and A. K. Bera (1980): "Efficient Tests for Normality, Homoscedasticity and serial Independence of Regression Residuals," *Economics Letters*, 6, 225–259.

——— (1987): "A Test for Normality of Observations and Regression Residuals," *International Statistical Review*, 55, 163–172.

Kolmogorov, A. (1933): "Sulla determinazione empirica di una legge di distribuzione," *G. Inst. Ital. Attuari*, 4, 83–91.

Smirnov, N. (1948): "Tables for estimating the goodness of fit of empirical distributions," *Annals of Mathematical Statistics*, 19, 279.

Stein, C. (1972): "A Bound for the Error in the normal approximation to the distribution of a Sum of Dependent Random Variables," in *Proc. Sixth Berkeley Symposium on Mathematical Statistics and Probability*.

Van der Vaart, A. W. (2007): *Asymptotic Statistics.* Cambridge University Press.

Appendix

k	$E(Y^k)$, $Y \sim \mathcal{N}(\mu, \sigma^2)$	$E(Z^k)$, $Z \sim \mathcal{N}(0,1)$
1	μ	0
2	$\mu^2 + \sigma^2$	1
3	$\mu^3 + 3\mu\sigma^2$	0
4	$\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$	3
5	$\mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4$	0
6	$\mu^6 + 15\mu^4\sigma^2 + 45\mu^2\sigma^4 + 15\sigma^6$	15
7	$\mu^7 + 21\mu^5\sigma^2 + 105\mu^3\sigma^4 + 105\mu\sigma^6$	0
8	$\mu^8 + 28\mu^6\sigma^2 + 210\mu^4\sigma^4 + 420\mu^2\sigma^6 + 105\sigma^8$	105

Table 1: Population moments of the $\mathcal{N}(0, 1)$ distribution

n	25	100	500	1000	5000
JB	0.0242	0.0398	0.0460	0.0477	0.0499
$H_{\{3\}}$	0.0554	0.0546	0.0512	0.0511	0.0521
$H_{\{4\}}$	0.0385	0.0428	0.0463	0.0474	0.0508
$H_{\{5\}}$	0.0220	0.0348	0.0457	0.0467	0.0513
$H_{\{6\}}$	0.0132	0.0196	0.0312	0.0358	0.0438
$H_{\{3,4\}}$	0.0541	0.0570	0.0536	0.0526	0.0513
$H_{\{5,6\}}$	0.0189	0.0287	0.0424	0.0462	0.0505
$H_{\{3,4,5,6\}}$	0.0546	0.0561	0.0592	0.0601	0.0583

Table 2: Rejection of true H_0

n	25	100	500	1000	5000
JB	0.5927	1.0000	1.0000	1.0000	1.0000
$H_{\{3\}}$	0.6276	0.9270	1.0000	1.0000	1.0000
$H_{\{4\}}$	0.6834	0.9842	1.0000	1.0000	1.0000
$H_{\{5\}}$	0.6232	0.9754	1.0000	1.0000	1.0000
$H_{\{3,4\}}$	0.6851	0.9810	1.0000	1.0000	1.0000
$H_{\{5,6\}}$	0.6467	0.9877	1.0000	1.0000	1.0000
$H_{\{2,3,4,5,6\}}$	0.7263	0.9944	1.0000	1.0000	1.0000

Table 3: Rejection of false H_0 (exp)

n	25	100	500	1000	5000
JB	0.0421	0.1033	0.2345	0.3594	0.8955
$H_{\{3\}}$	0.1122	0.1376	0.1549	0.1627	0.1684
$H_{\{4\}}$	0.0942	0.1650	0.3563	0.5223	0.9708
$H_{\{5\}}$	0.0631	0.1261	0.2372	0.2920	0.4016
$H_{\{6\}}$	0.0437	0.0892	0.1956	0.2687	0.5231
$H_{\{3,4\}}$	0.1189	0.1892	0.3550	0.5001	0.9567
$H_{\{5,6\}}$	0.0592	0.1193	0.2470	0.3293	0.5842
$H_{\{3,4,5,6\}}$	0.11045	0.1857	0.3694	0.5146	0.9475

Table 4: Rejection of false H_0 (t, df=25)

n	25	100	500	1000	5000
JB	0.0338	0.0751	0.1359	0.1902	0.5344
$H_{\{3\}}$	0.0883	0.1026	0.1077	0.1091	0.1124
$H_{\{4\}}$	0.0694	0.1089	0.1925	0.2710	0.6957
$H_{\{5\}}$	0.0459	0.0854	0.1460	0.1733	0.2356
$H_{\{6\}}$	0.0288	0.0561	0.1124	0.1470	0.2561
$H_{\{3,4\}}$	0.0925	0.1303	0.1989	0.2648	0.6457
$H_{\{5,6\}}$	0.0413	0.0770	0.1464	0.1861	0.3080
$H_{\{3,4,5,6\}}$	0.0837	0.1265	0.2156	0.2855	0.6307

Table 5: Rejection of false H_0 (t, df=40)

n	25	100	500	1000	5000
JB	0.0029	0.0112	0.9168	1.0000	1.0000
$H_{\{3\}}$	0.0097	0.0068	0.0065	0.0070	0.0069
$H_{\{4\}}$	0.0234	0.1743	0.9508	0.9998	1.0000
$H_{\{5\}}$	0.0026	0.0026	0.0021	0.0022	0.0021
$H_{\{6\}}$	0.00595	0.0796	0.8518	0.9977	1.0000
$H_{\{3,4\}}$	0.0103	0.0702	0.8594	0.9988	1.0000
$H_{\{5,6\}}$	0.0012	0.0205	0.6375	0.9832	1.0000
$H_{\{3,4,5,6\}}$	0.0101	0.0897	0.9017	0.9995	1.0000

Table 6: Rejection of false H_0 (Bimodal)

n	25	100	500	1000	5000
JB	0.9566	0.9991	0.9801	0.8156	0.2347
$H_{\{3\}}$	0.9999	0.9975	0.9914	0.5186	0.2971
$H_{\{4\}}$	1.0000	1.0000	0.9984	0.8808	0.2687
$H_{\{5\}}$	1.0000	0.9999	0.9996	0.9981	0.7230
$H_{\{6\}}$	1.0000	1.0000	0.9984	0.8808	0.2687
$H_{\{3,4\}}$	1.0000	1.0000	0.9987	0.8720	0.2474
$H_{\{5,6\}}$	1.0000	1.0000	1.0000	1.0000	0.9973
$H_{\{3,4,5,6\}}$	1.0000	1.0000	1.0000	1.0000	0.9980

Table 7: Rejection of true H_0, outlier ($x = 5$) added

Auction Theory and Applications

Prof. Dr. Benny Moldovanu*†

Introduction

In an auction participants (repeatedly) submit bids representing their demand or supply functions. Then, accepted trades and transaction prices are calculated by some explicit aggregation procedure. Auctions directly implement the ideas implicit in the Walrasian analysis of finding prices that are consistent with individual maximization and that equate demand and supply. In contrast to general equilibrium analysis, auction theory is based on the premise of individual strategic behavior. It offers explicit models of price formation and allocative distribution that can be applied also to small markets. Auctions have been continuously used since antiquity, and they remain simple and ubiquitous means of conducting multilateral trade. The belief that auctions yield competitive outcomes even if information is dispersed is behind the practical appeal of auctions, and behind their recent popularity.

Modern, large-scale applications include treasury auctions, spectrum auctions, private and government procurement auctions, CtoC, BtoC and BtoB internet

*Benny Moldovanu has been a full professor at the University of Bonn since 2002. Further, he was visiting professor at Yale University in 2004, at the University College London in 2005 and at Northwestern University in 2009. The winner of the 2010 Advanced Investigators Grant, awarded by the European Research Council, is one of the leading researchers in the field of mechanism design and has published in the leading international academic journals. Moldovanu is member of CEPR and of the European Economic Association and has designed several auctions for large government and private sector projects.

†The present material is based on insights found in Jehiel and Moldovanu (2006) and Jehiel and Moldovanu (2003). I am grateful to Philippe Jehiel for many years of fruitful collaboration.

auctions, real-estate auctions, commodities auctions and asset auctions following bankruptcy. In practice auctions are often used in order to achieve other important goals, besides, or instead efficiency: 1) Revenue-maximization. The seller of an item at eBay, say, does not care much about allocative efficiency, but rather aims at maximizing the price he gets. 2) Information Aggregation and Revelation. Auctions aggregate bids which are made based on the basis of privately known demand and supply functions. Hence, the resulting prices can also be seen as aggregators of private information. For example, data collected at auctions of liquidity organized by central banks is an important indicator for future monetary policy. 3) Transparency and Speed. In serious auctions the rules are precise, fixed in advance, and applied equally to all participants. Auctions provide the means for achieving quick, frictionless transactions. Speed is particularly important for perishable goods such as fish, vegetables or flowers. Accordingly, most whole-sale markets for agricultural products have long been organized as auctions.

The above goals are not independent of each other. Depending on the economic environment and on the market process, there are subtle relations among them, and sometimes even severe conflicts.

Traditionally, the focus of auction theory has been on models that view auctions as isolated events. In practice, however, auctions are often part of larger transactions: for example, in privatization exercises such as license allocation schemes auctions shape the size and composition of future markets. Thus the auction typically affects the nature of the post-auction interaction among bidders. On the other hand, anticipated scenarios about the future interaction influence bidding behavior: already at the bidding stage agents need to care about who gets what, and about the information revealed to, or possessed by others, since these features will be reflected in the equilibrium of the post-auction interaction. Thus allocative and informational externalities naturally arise in models that embed auctions in larger economic contexts.

General equilibrium analysis has identified several forms of externalities as obstacles on the road towards economic efficiency. The First Welfare Theorem fails in the presence of allocative externalities, i.e., when agents care about the physical consumption bundles of others. Akerlof's famous analysis demonstrated that the First Welfare Theorem may also fail when agents care about the information held by others. Thus, given the obstacles created by externalities in general equilibrium, it is of interest to understand what are the parallel consequences of external effects in auctions. This has been one of my main research topics (in cooperation with Philippe Jehiel and other colleagues).

An Illustration

As an illustration, consider the European process of allocating UMTS spectrum and licenses to telecom firms in $2000 - 2001$. The auctioned objects were licenses to operate a third-generation mobile telephony network in a certain country. Since per-firm industry profit in oligopoly decreases in the number of active firms, incumbents were also driven by entry preemption motives (e.g., the need to avoid further losses relative to the status quo) which translate into increased willingness to pay for licenses and capacity. To see this force at work, let us recall the German experience. Bidders did not directly submit bids for licenses but, instead, on 12 blocks of spectrum. A bidder obtained a license only if he acquired at least two blocks, and a bidder was allowed to acquire three blocks. Thus, the number of licensed firms could vary between 0 and 6. There were 7 bidders, including 4 GSM incumbents. The auction lasted for 173 rounds, and the winning firms were the 4 incumbents and two new entrants. Each licensed firm acquired 2 blocks and each license cost approximately Euro 8.4 Bn (4.2 Bn per block).

The most interesting thing occurred after one of the potential entrants left the auction in round 125, after the price reached Euro 2.5 Bn per block. Since 6 firms were left bidding for a maximum of 6 licenses, the auction could have stopped **immediately**. Instead, bidding in order to acquire more capacity and to reduce the number of competitors continued until round 173 (only intense pressure from stock markets and bond rating agencies stopped it). Compared to round 125, there was **no change in the physical allocation** but firms where, collectively, Euro 20 Bn **poorer**! Given the immense sum that was paid for licenses, the two new entrants went bankrupt (this is called a "winner's curse" in auction jargon) and did not build networks, thus leaving Germany with the four old network operators.

Single-Object Auctions

The simplest theoretical setup is one in which bidders know how much they value the good for sale, but are uncertain as to how much other bidders value the good. This is the so-called "private value" paradigm. In the ascending price (or English) auction the price gradually increases, bidders may drop out at any point in time, and the auction stops when there is only one bidder left. The last active bidder buys the good at the price where the auction stopped. In the private value setting, the ascending price auction induces an efficient outcome The reason is that it is a dominant strategy to drop out whenever the price reaches one own's valuation. In a Nobel-prize winning paper, Vickrey (1961) proposed a condensed version of this auction, now called the Vickrey auction or, in the context of one-object auctions, the sealed-bid second-price auction, in which agents secretly place bids. The bidder with the highest bid wins the object and pays the second-highest bid. Vickrey observed that in the second-price sealed-bid auction it is a dominant strategy to bid one own's valuation. Hence this format is here equivalent to the ascending price auction. It turns out that these formats, augmented by a reserve price are also **revenue maximizing** whenever bidders' signals about their values are independent of each other, agents are risk neutral, ex-ante symmetric, and the seller is bound to sell her good.

Laymen feel that more money can be extracted by requesting that the winner pays the highest bid, rather than the second highest bid. Such a format corresponds to the commonly used first-price sealed-bid auction. The argument ignores that bidders react to a change of format by adjusting their bidding strategies. In a first-price auction, a bidder who bids her own value never makes any money. Hence bidders bid less than their value, and bids will be lower than in the second-price version (but the seller receives the highest bid, not the second highest). It turns out that, in expected terms, the first-price auction generates exactly the same amount of revenue as the second-price auction as long as the agents are symmetric and risk neutral, and obtain independent signals about their respective values. This result is the celebrated Revenue Equivalence Theorem, first noticed by Vickrey. Many of the above strong conclusions heavily rely on ex-ante symmetry, risk-neutrality, the absence of budget constraints, signal independence, and the absence of informational or allocative externalities, and need to be adjusted if these features are not present!

The analysis of auctions with externalities is quite subtle even in the one-object case. One reason is that the notion of "valuation" is not well defined a-priori. Specifically, how much a bidder is ready to pay depends on his expectation about what will happen if he does not buy the object. If he expects the winner to be a tough competitor, he will be ready to pay a high price; if he expects the winner to be a soft competitor, his willingness to pay will be low.

Example 1. ***A Takeover Contest.*** *Consider three firms bidding for a fourth in a takeover contest. Consider the following expectations about future scenarios in the post-takeover industry: due to synergies, each bidding firm expects to make* *an extra profit of* π *if it wins the contest; if firm* 1 *wins, firm* 2 *expects a relative decrease in profits of* α *(and vice versa); firm* 3 *expects a decrease in profits of* $\gamma < \alpha$ *if firm* 1 *or firm* 2 *wins; finally, firms* 1 *and* 2 *expect to be unaffected if* 3 *wins. From the firms' viewpoint (e.g. abstracting from consumers' surplus), the efficient buyer is firm* 3 *(since the other firms do not expect to suffer a loss in that scenario). Consider a first price sealed-bid auction: In one equilibrium, firm* 3 *indeed wins by bidding only* π*. But in another, firms* 1 *and* 2*, who are very afraid of each other, engage in a race and one of them wins by bidding up to* $\pi + \alpha$*. In that case, both firms suffer a loss of* $-\alpha$*! If firms* 1 *and* 2 *think that the expensive race is going to happen because they cannot coordinate to let* 3 *win, they will have incentives to commit not to participate at the auction in the first place. For example, if firm* 1 *withdraws,* 3 *necessarily wins since it is willing to bid up to* $\pi + \gamma$*, while firm* 2 *is willing to bid only up to* π *(since* 1 *poses no danger anymore). This scenario is, in fact, better for firm* 1 *than participating in the race with* 2.

If, as above, the allocative externalities are due to market structure considerations, the notion of economic efficiency should not be solely based on considerations about firms' welfare. Instead, the welfare considerations should also include the consumer's surplus in each possible future scenario. But, a more thoughtful design that addresses the consumers' interest may generate low revenue.

Example 2. ***Competition over Monopoly Rents.*** *There are two licenses* A *and* B *for sale. There are two firms* $i = 1, 2$ *competing for the two licenses. Each firm* i *is allowed to buy both* A *and* B*. If firms* 1 *and* 2 *each buy one license, price-competition is assumed to drive profits down significantly (say to zero). If firm* i *buys both licenses, it earns monopoly profits* π_i*. We assume that* $\pi_1 > \pi_2$*. Consider the standard format where firms simultaneously submit two bids* b_i^A, b_i^B *for licenses* A *and* B*, respectively. Each license is allocated to the highest bidder* *on that license, who pays the bid. Since one license is worthless (as long as the other is also sold), the outcome of this auction is that firm* 1 *gets the two licenses and pays* π_2 *for it. In other words, the auction selects a monopoly structure, and the resulting market structure is not desirable from the consumers' viewpoint. If the government allows each firm to buy only one license, a duopoly may emerge (which is presumably better for consumers and total welfare) but the auction's revenue will be low.*

Multi-object auctions

Multi-object auctions raise a large number of difficulties. Even when externalities are absent, multi-unit demand, heterogeneity and complementarities induce complex demand functions, which are difficult to map in reasonably simple auction formats. Simple formats necessarily restrict bidders in some dimension, creating complex strategic effects that affect the auction's performance. We illustrate below several difficulties arising in such auctions.

With single unit demand bidders and with k homogeneous units, the $k+1^{th}$ -price auction induces an efficient allocation. When bidders have multi-unit demand, a seemingly easy generalization of this format is the uniform-price auction: bidders submit demand curves (i.e., bids for 1 up to k units), and the units are allocated to maximize the values expressed by the submitted demand curves; every allocated unit is sold at the same minimum price where aggregate demand coincides with the number of supplied units k.

Unfortunately, if bidders have multi-unit demand, the uniform price auction may lead to an inefficient allocation since bidders have an incentive to lower their demand on all units (but the first). In doing so, they affect downwards the selling price and pay a lower price on the remaining units.

Example 3. ***Demand Reduction.*** *Three identical cases of Bordeaux wine are sold through a uniform price auction. There are three potential bidders,* $i = 1, 2, 3$. *Bidders* 2 *and* 3 *are interested in one case only; their valuations are* 1 *and* 0.25*, respectively. Bidder* 1 *is potentially interested in all three cases. His valuation is* 10 *for the first unit,* 5 *for the second, and* 2 *for the third. Efficiency in the above example dictates that bidder 1 gets all three cases. Bidders 2 and 3 have a dominant strategy—to bid their values for one case. If bidder* 1 *expresses his true demand, the minimum price where demand equals supply is (slightly above)* 1*, and bidder 1's payoff is given by* $10 + 5 + 2 - 3 = 14$. *However, in the equilibrium of the uniform price auction, bidder 1 will decrease his demand to only two cases (e.g., he will bid* 15 *for either two or three units). This lowers the selling price per unit to* 0.25*, yielding for bidder 1 a payoff of* $10 + 5 - 0.5 = 14.5$. *The allocation is inefficient since bidder 2 obtains one case.*

Complementarities are particularly troublesome in auction formats where bids can be placed only on individual objects, but not on bundles (or packages). Bids cannot then fully reflect the magnitude of complementarity, creating inefficiencies. The theoretically correct way to deal with complementarities is to allow for "combinatorial" auctions where agents can place bids directly on bundles. Forbidding such bids may give rise to the so called *exposure problem:*

Example 4. ***The Exposure Problem.*** *There are two parking slots, and two bidders. Bidder 1 has a car and a trailer, and he values the two parking slots together at $100, while attaching a value of zero to each individual slot. Bidder 2 has only a car and values any slot at $75. The value of two slots is also $75 for this bidder. Efficiency dictates that bidder 1 gets both slots. But if the auction format does not allow bids on the whole package of two slots, any positive bid on an individual slot exposes bidder 1 to the danger of obtaining only that slot alone—an alternative valued at zero. The only equilibrium in a simultaneous* *ascending auction without combinatorial bids is for the first bidder to drastically reduce demand by non-participation (since otherwise he needs to bid up to $75 per slot in order to outbid the other player). Bidder 2 inefficiently wins a slot by placing a minimum bid on it. In the presence of incomplete information, bidder 1 will bid only if he attaches a sufficiently high probability to the event in which bidder 2 has a low valuation. If bidder 1 decides to bid (based on his information), and if it turns out that bidder 2 has a high value, bidder 1 will regret his decision.*

The main problem with combinatorial auctions is that they may be very complex to conduct and participate at. Besides the structural complexity arising in large auctions, combinatorial bids also induce some subtle strategic problems, such as inefficient free riding.

Example 5. ***Free Riding.*** *Two regional radio licenses are put for sale. There are three potential bidders* $0, 1, 2$. *Bidder* 0, *who needs national coverage, values only the bundle* $\{1, 2\}$ *at* v^{12}. *Bidder* i, $i = 1, 2$, *values only license* i *at* v^i. *Assume that* $v^1 + v^2 > v^{12}$. *Each bidder* i *simultaneously submits a bid* b_i *for whatever good or bundle she wishes. The goods are allocated so as to maximize the revenue generated by the bids, and each bidder pays for the goods he receives according to the bid he submitted ("pay-your-bid" auctions). Efficiency dictates that bidder* i *receives object* i, $i = 1, 2$. *But in equilibrium there will be a "war of attrition" between bidders* 1 *and* 2. *Instead of bidding up to* v^1 *on object* 1, *bidder* 1 *prefers to place a low bid on object* 1 *(say* $v^{12} - v^2$*), hoping that bidder* 2 *will make a high bid on object* 2 *(say* v^2*). Similarly, bidder* 2 *prefers to place low bid (say* $v^{12} - v^1$*), hoping that bidder* 1 *will make a high bid on object* 1 *(say* v^1*). As a consequence, there is an equilibrium in mixed strategies where bidder* 0 *gets the bundle with positive probability.*

In one-object symmetric settings, standard auctions are efficient and revenue-maximizing (at least if the seller is not allowed to retain the good). Thus, efficiency and revenue go hand in hand. This ceases to be true in multi-object settings, even if the situation is symmetric and there are no complementarities.

Example 6. ***Efficiency or Revenue?*** *Consider an auction for two objects A and B, and two bidders,* 1 *and* 2*. For both agents, the valuations for the bundle* $\{A, B\}$ *are given by the sum of the valuations for the individual objects, and assume these to be as follows:*

$$\begin{aligned} v_1^A &= 10; v_1^B = 7 \\ v_2^A &= 8; v_2^B = 12 \end{aligned}$$

The value maximizing auction (which puts the objects in the hand of those who value them most) is simply given by two separate second-price auctions, one for each object. Then object A goes to bidder 1 *for a price of* 8*, while object B goes to bidder* 2 *for a price of* 7*. Total revenue is* 15*. But, consider now a single second-price auction for the entire bundle* $\{A, B\}$*. Then the bundle will be acquired by bidder B, for a price of* 17 *! Hence, revenue is higher in the bundle auction, but object A is mis-allocated.*

The presence of multiple objects for sale creates many new possibilities for “tacit” collusion in which firms coordinate their bids instead of competing. The main idea is that it may be preferable to share the objects at low prices instead of trying to buy more of them while pushing prices up. Ascending price formats are more vulnerable to such behavior since they offer repeated opportunities to signal intentions and future behavior. In contrast, sealed bids greatly reduce the scope of signalling, but run the risk of yielding an inefficient allocation since private information does not get properly aggregated.

Example 7. ***The German GSM Auction.*** *In October 1999 Germany auctioned 10 additional blocks of paired spectrum to the four GSM incumbents. Nine blocks were identical, each consisting of* 2×1 *MHz, while the tenth block consisted of* 2×1.4 *MHz. After the first round, the high bidder on all 10 blocks was Mannesmann (one of the two large players), which offered DM 36.360.000 for each of blocks 1-5, DM 40.000.000 for each of the blocks 6-9 (which, recall, were* ***identical*** *to blocks 1-5), and DM 56.000.000 for the larger block 10. In the second round, T-Mobil (the other big player) bid DM 40.010.000 on blocks 1-5, and the auction closed! Hence, each of the two larger firms got 5 blocks , at a price of DM 20.000.000 per MHz. Here is what one of T-Mobil's managers said: "No, there were no agreements with Mannesmann. But Mannesmann's first bid was a clear offer. Given Game Theory, it was expected that they show what they want most."*

Conclusion

Each market process creates specific strategic incentives for the participants and leads to specific distortions. Thus, the "rules of the game" do matter and deserve attention. Different auction formats lead to unavoidable trade-offs among goals. Thus, a clear determination of the auction's goals is indispensable before the choice of the auction format can be discussed. The precise structuring of traded goods and feasible bids play a main role in determining whether the auction procedure can accurately represent the agents' preferences and lead to a desired outcome.

Allocative or informational externalities, multi-unit demand, heterogeneity and complementarities induce complex demand or supply functions, which are difficult to map in reasonably simple auction formats. Such formats necessarily restrict bidders in some aspects, creating complex strategic effects that affect the auction's performance. If the auction's allocation influences some future interaction among bidders, these will take this effect into account at the bidding stage. Thus, the future interaction also influences the auction's outcome through the participants' expectations. In complex environments serious design calls for an integrated approach that combines the insights of Auction Theory with the traditional concerns of regulation and competition policy.

References

Jehiel, P., and B. Moldovanu (2003): "An Economic Perspective on Auctions," *Economic Policy*, 18(36), 271–308.

——— (2006): "Allocative and Informational Externalities in Auctions and Related Mechanisms," in *The Proceedings of the 9th World Congress of the Econometric Society*, ed. by T. P. Richard Blundell, Whitney Newey. Cambridge University Press.

Klemperer, P. (1999): "Auction Theory: A Guide to the Literature," *Journal of Economic Surveys*, 13(3), 227–286.

Krishna, V. (2002): *Auction Theory*. Academic Press.

McAfee, R. P., and J. McMillan (1987): "Auctions and Bidding," *Journal of Economic Literature*, 25(2), 699–738.

Milgrom, P. R. (1987): "Auction Theory," in *Advances in Economic Theory: Fifth World Congress*. Cambridge University Press.

——— (2004): *Putting Auction Theory to Work*. Cambridge University Press.

Vickrey, W. (1961): "Counterspeculation, Auctions and Competetive Sealed Tenders," *Journal of Finance*, 16(1), 8–37.

Wilson, R. (1993): "Strategic Analysis of Auctions," in *Handbook of Game Theory*, ed. by S. H. R. Aumann. Elsevier.

Wolfstetter, E. (1996): "Auctions: An Introduction," *Journal of Economic Surveys*, 10(4).

Common Pools — Why a European Fiscal Union will Make Things Worse

Prof. Dr. Jürgen von Hagen*

The root of excessive deficits and debts is in the lack of proper governance over common pool problems of public finance (Kontopoulos and Perotti (1999), von Hagen and Harden (1995), Weingast, Shepsle, and Johnsen (1981), Wyplosz and Kostrup (2010), Hallerberg, Strauch, and von Hagen (2009)). The common pool problem of public finance is the result of financing public policies targeted at specific groups in society from a general tax fund, which creates an externality: Those enjoying the marginal benefit from an extra euro of public spending are not those bearing the marginal cost of funding it. If they did, they would choose the level of spending that equates the marginal benefit and cost of funding. But since they generally do not, those benefitting from a policy tend to ask for higher levels of spending, deficits, and debts.

This common pool problem of public finances manifests itself in a number of

* Jürgen von Hagen is Professor of Econcomics at the University of Bonn and Director of the Institute for International Economic Policy. He earned his PhD in economics at the University of Bonn in 1985 whereto he returned in 1996 as professor of economics after teaching assignments at Indiana University, USA, and at the University of Mannheim, Germany. In 1997, he became the first Winner of the Gossen Prize. Von Hagen is a member of CEPR, of the Academic Advisory Council of the German Federal Ministry of Economics and director of the Institute for International Economic Policy. His main research fields are international and monetary macroeconomics as well as public finance, where he has published numerous articles in leading international academic journals. Amongst others, von Hagen has been a consultant to the IMF, the European Commission, the Federal Reserve Board, the European Central Bank, and the World Bank.

ways. The first concerns the level of public spending and taxation. Representatives of different political constituencies compete for financial resources and must reach a decision on the level of taxation and spending and the distribution of spending over a range of public policies (von Hagen and Harden (1995), Kontopoulos and Perotti (1999)). The more narrowly individual policies are targeted toward individual constituencies, the more pervasive the common pool problem becomes.[1] Cultural, ethnic, and other divides among the population aggravate the common pool problem, since each constituency pays less attention to the fiscal burdens falling on the other.

A second manifestation of the common pool problem occurs when current government spending can be financed by borrowing, since this gives today's decision makers access to future general tax funds. The result is excessive borrowing compared to a situation in which the common pool externality is fully internalized (see von Hagen and Harden (1995), Wyplosz and Kostrup (2010)).

A third manifestation concerns the financial relations between different levels of government, where the degree of vertical imbalance, i.e., the ratio of spending at the lower level to the own tax revenue collected by lower-level units, is a critical parameter. The greater the degree of vertical imbalance, the more the sub-central units depend on revenues transferred from the central government. Such transfers give the lower units the opportunity to spend taxes collected from citizens in other parts of the federation or country. They invite strategic behavior to extract more transfers from the higher-level government (see e.g. Careaga and Weingast (2000)). If the lower units can borrow from banks or capital markets, bailouts of over-indebted jurisdictions are a particularly pernicious form of vertical transfers and common pool problems (see Rodden (2003), von Hagen, Bordignon, Grewal, Peterson, and Seitz (2000)). Such bailouts have been the cause of fiscal and

[1] The classical analysis of this problem is Weingast, Shepsle, and Johnsen (1981).

currency crises in weak federations such as Argentina and Brazil in the 1980s and 1990s and a recurrent problem in Germany since the late 1980s (see von Hagen, Bordignon, Grewal, Peterson, and Seitz (2000)).

A fourth manifestation of the common pool problem is in the event of a *war of attrition* (Alesina and Drazen (1991)). This is a situation requiring a large fiscal adjustment, in which the representatives of some large constituencies in society cannot agree on the distribution of the adjustment burden, and public debt keeps increasing as no agreement is being reached. The current European debt crisis in some aspects resembles a *war of attrition* as the European governments seem unable to reach an agreement on how to distribute the cost of crisis resolution over the citizens of their countries. Northern Europeans are unwilling to accept this cost pointing at the past lack of reforms and fiscal adjustments in Southern Europe, while Southern Europeans argue that they cannot bear a larger burden than they already do.

A final manifestation of the common pool problem is the bailout of large financial institutions Europe witnessed in the financial crisis of 2008 – 2009. Governments loved shmoozing with bankers, because they project an image of economic relevance. As the bankers went under, they quickly invented the concept of systemic relevance for their institutions to prove the inevitability of being rescued by tax payers' money, and the politicians willingly gave in. All of these manifestations have played a role in the emergence of the public debt crisis in Europe.

The canonical common pool model considers a government consisting of a group of decisionmakers all drawing money from the same general tax fund to finance projects benefitting their own constituencies. Current spending can be paid for with current or future tax revenues, i.e., the government can borrow at a given interest rate. Spending on each project has positive and declining marginal benefit, while taxation has positive and increasing marginal cost. Each decisionmaker takes into account only the share of the marginal cost that falls on his constituency. The decisionmakers play a non-cooperative game in which each submits a bid specifying an amount of spending taking all other bids as given. The (symmetric) equilibrium of this game is characterized by inefficiently high levels of spending on each project and too large deficits relative to a solution that would maximize social welfare (von Hagen and Harden (1995)). Theory and ample empirical research shows that this spending and deficit bias increases with the number of decisionmakers and constituencies involved, the number and the depth of political, cultural, and ethnic cleavages in society, and the degree of opacity of the decisionmaking process. The difference between the collectively and the individually optimal solutions, which is at the heart of the externality problem, implies that, even if an agreement to adopt the former could be reached, each individual decisionmaker has a strong incentive to deviate from it and secretly increase spending in his area. Implementing such an agreement, therefore, needs strong enforcement.

The deficit bias can be eliminated by imposing a cost on each decisionmaker for contributing to a deficit by increasing spending in his domain. To reach the socially optimal spending and deficit levels, this cost must be sufficiently large, and the larger the number of decisionmakers involved, the larger must be the cost of contributing to the deficit to enforce the socially optimal policy.

Since the beginning of the public debt crisis in Europe, politicians, including many in Germany, have been calling for a European Fiscal Union to complement the European Monetary Union and ensure the sustainability of public finances of its member states. Implicitly, the claim is that collective decisions over public sector deficits and debts at the European level would be a better and safer way to implement what is socially optimal than decentralized decisions at the national level. What does the above argument about the common pool problem imply for the proposed European Fiscal Union?

Such a Union would create a European tax fund from which spending in individual member states would be financed. Furthermore, the Fiscal Union would be vested with the authority to borrow on behalf of the member states. Compared to the setting in each member state, a European Fiscal Union would imply a larger number of constituencies and their representatives drawing from the common tax fund, and more and deeper cleavages as different nationalities, cultures, religions, regional identities etc. would come into play. Experience with other European decisionmaking processes suggests that it would also lead to more opacity and less democratic control of the decisionmaking processes. Therefore, a European Fiscal Union can only aggravate the common pool problem. It follows that a European Fiscal Union has an even greater deficit bias than each national fiscal policy. It would make the European public debt problem worse.

One might argue that the decisions in a European Fiscal Union could be subject to rules enforcing small deficits and that decisionmakers contributing to larger deficits would be punished for their behavior. Our analysis, however, implies that the punishment would have to be stronger than the punishment in the case of national fiscal policies. It is not clear, why and how that should come about. Experience with the European fiscal framework so far has shown that enforcement of the common rules is weak. National administrations have ample room for creative accounting and hiding deficit financing.[2] If anything the political cost of causing large deficits would be smaller at the European than at the national level.

Alternatively, a European Fiscal Union could come with the restriction that tax revenues collected in a given member country can only be spent on projects

[2] On creative accounting in the euro area see von Hagen and Wolff (2006) For using private institutions to create public debt recall that the hidden inflation in Germany in the late 1930s was caused by the issuing of seemingly private trade bills (so-called Mefo Wechsel) by the German government.

benefitting constituencies in the same country. The Union's common fiscal policy would then be a collection of national fiscal policies with the same properties as under a decentralized arrangement. Obviously, this would simply replicate the national policies with the same deficit bias.

Proponents of the Fiscal Union argue that the Stability and Growth Pact and the new Fiscal Compact (see European Commission (2012)) in the European Monetary Union will yield an improvement over purely national fiscal policies nevertheless. The idea is that they create a process of monitoring national fiscal policies and exert pressures on national governments to reduce the spending and deficit bias. In practice, however, they do not serve this function. National governments merely submit their budget plans to the European Commission for approval. The Commission makes its assessment of them based on purely statistical criteria which have neither economic nor political content. And again, it lacks the means to enforce the European rules effectively. At the same time, however, the Stability and Growth Pact and the new Fiscal Compact create the illusion of a European framework assuring fiscal discipline. It dilutes accountability at the national level, as policymakers can point to the adherence to purely formal criteria to excuse excessive spending and borrowing at the national level. Furthermore, the complexity of the European procedures adds opacity to the decisionmaking processes. No improvement over national policies can be expected.

I summarize these arguments in the following *Decentralization Theorem*: With respect to excessive spending and deficits, neither a European Fiscal Union nor the Stability and Growth Pact *cum* Fiscal Compact can do better than national fiscal policies. Chances are, it will do much worse.

One might argue, of course, that there are other advantages of a common fiscal policy. Coordinating anti-cyclical policies is a traditional example (see von Hagen and Mundschenk (2003)). But it is unclear that the externalities caused by spill-overs of aggregate demand from one country to another are truly significant. De Grauwe (2011) argues that a common fiscal policy could implement transfer payments to countries hit by negative economic shocks and act as an insurance mechanism against asymmetric shocks. This is true, but empirical evidence from existing large federations suggests that the need for such insurance is not very large. In Germany, for example, the system of horizontal transfers of tax revenues among the states and vertical transfers between the federal government and the state governments has only spurious insurance capacity against asymmetric shocks. It is mainly a redistributive mechanism (see Hepp and von Hagen (2012a), Hepp and von Hagen (2012b)). Overcoming collective action problems in the face of a crisis is another argument. The claim here is that a common policy would be able to react faster to a crisis than coordinated national policies. Obviously such a claim is unsubstantiated unless the decision procedures at the national and the European level are clearly defined. After all, it is also popular to say that the ECB has been reacting too slowly to the crisis, and the ECB represents a truly common policy. Whatever the argument in favor of a common fiscal policy is, it has to be weighed against the perilious consequences of a substantially larger common pool problem.

References

Alesina, A., and A. Drazen (1991): "Why are Stabilizations Delayed?," *American Economic Review*, 81:5, 1170–1181.

Alesina, A., and R. Perrotti (1996): "Fiscal Discipline and the Budget Process," *American Economic Review*, 86, 401–407.

Careaga, M., and B. R. Weingast (2000): "The Fiscal Pact With the Devil: A Positive Approach to Fiscal Federalism, Revenue Sharing, and Good Governance," Working Paper, Stanford University.

De Grauwe, P. (2011): "The Governance of a Fragile Eurozone," CEPS Policy Brief 346.

European Commission (2012): "Six-pack? Two-pack? Fiscal compact? A short guide to the new EU fiscal governance," Available online at http://ec.europa.eu/economy_finance/articles/governance/2012-03-14_six_pack_en.htm. Accessed November 12, 2012.

Hallerberg, M., R. R. Strauch, and J. von Hagen (2009): *Fiscal Governance in Europe*. Cambridge University Press.

Hepp, R., and J. von Hagen (2012a): "Fiscal Federalism in Germany: Stabilization and Redistribution Before and After Unification," *Publius Journal of Federalism*, 42:2, 243–259.

——— (2012b): "Regional Risk Sharing in Germany: 1970-2006," Oxford Economic Papers in print.

Kontopoulos, Y., and R. Perotti (1999): "Fiscal Institutions and Fiscal Performance," in *Government Fragmentation and Fiscal Policy Outcomes: Evidence from OECD Countries*, ed. by J. Poterba, and J. von Hagen. University of Chicago Press.

Rodden, J. (2003): "Fiscal Decentralization and the Challenge of Hard Budget Constraints," *Cambridge: MIT Press.*

Velasco, A. (1999): "Fiscal Institutions and Fiscal Performance," in *Debts and Deficits with Fragmented Fiscal Policymaking*, ed. by J. Poterba, and J. von Hagen. University of Chicago Press.

von Hagen, J., J. M. Bordignon, B. Grewal, P. Peterson, and H. Seitz (2000): "Subnational Government Bailouts in OECD Countries: Four Case Studies," *Interamerican Development Bank Research Network Working Paper R-399.*

von Hagen, J., and I. J. Harden (1995): "Budget Processes and Commitment to Fiscal Discipline," *European Economic Review*, 39, 771–779.

von Hagen, J., and S. Mundschenk (2003): "Fiscal and Monetary Policy Coordination in EMU," *Journal of International Finance and Economics*, 8, 279–295.

von Hagen, J., and G. Wolff (2006): "What Do Deficits Tell Us About Debt? Empirical Evidence on Creative Accounting With Fiscal Rules in the EU," *Journal of Banking and Finance*, 30:12, 3259–3279.

Weingast, B. R., K. A. Shepsle, and C. Johnsen (1981): "The Political Economy of Benefits and Costs: A Neoclassical Approach to Distributive Politics," *Journal of Political Economy*, 89(41), 642–664.

Wyplosz, C., and S. Kostrup (2010): "A common pool theory of supranational deficit ceilings," *European Economic Review*, 54:2, 269–278.

Ratings of Sovereign Debt during the Euro Crisis — An Empirical Assessment

Markus Behn, Jonas Sobott, Rüdiger Weber, Dorje Wulf*

Introduction

While valid estimates of securities' default risk are a crucial ingredient for making appropriate investment decisions, the course of the European debt crisis has shown that it is difficult or even impossible to obtain such estimates for sovereign bonds. Rating agencies, which investors trusted to make informed judgments about a country's creditworthiness in the past, have been criticized for systematically underestimating default risk and misrating several European countries and securities prior to the crisis. By contrast, during the crisis several politicians and bureaucrats complained about downgrades that were−in their eyes−not justified by macroeconomic fundamentals. A recent example is given by Christian Noyer, the governor of the French central bank, who claimed that rating agencies have become "incomprehensible and irrational" in their assessments (Deen (2011)).

* Markus Behn, Jonas Sobott, Rüdiger Weber and Dorje Wulf currently are students of Economics (M. Sc.) at the University of Bonn. Together with their instructor, Prof. Dr. Rainer Haselmann, they won the first place of this year's Postbank Finance Award which is endowed with 50.000 Euro for this paper on the meaningfullness of country ratings for financial investments. A total of 28 student teams from 25 universities and colleges in Germany and Austria participated. Postbank Finance Award has been granted annually since 2003. The aim is to promote innovative and scientifically sound answers to current financial and economic issues.

In our paper "Welche Aussagekraft haben Länderratings? Eine empirische Modellierung der Ratingvergabe während der europäischen Staatsschuldenkrise", we address this issue by examining the informational content as well as the reliability of country ratings for investment decisions. The aim of the paper is to show potential inconsistencies in the rating process of European countries in the period between 1995 and 2011.

Like other studies before us we first show that it is possible to replicate country ratings with a few—publicly available—macroeconomic variables.[1] Our simple model is able to explain more than 90% of the variation in country ratings and questions the necessity of agencies' complex and nontransparent models. In a second step we show that the weight that rating agencies attribute to the individual macroeconomic factors is not consistent over time. In particular, the overall debt level of a country has a much greater impact on the rating since the beginning of the European debt crisis. Rating agencies seem to adjust the rating process in order to be in line with public perception, which questions the usefulness of ratings for long-term investment decisions. Our third result points into the same direction: We show that pre-period prices of Credit Defaults Swaps (CDS) can be used to predict changes in country ratings, while the opposite relationship does not hold. In a way, this result could be expected as agencies aim to provide long-term assessments of a debtor's creditworthiness. However, one might ask the question about the added value of ratings if markets give corresponding signals before changes in ratings occur. As our final result we find no evidence of arbitrary downgrades of countries during the European debt crisis. All downgrades are justified by fundamentals. Hence, while agencies seem to adjust their

[1] For other studies that show that publicly available macroeconomic variables are often sufficient to replicate country ratings see e.g. Cantor and Packer (1996), Afonso (2003) or Butler and Fauver (2006). A study that is closely related to our own is the one by Gärtner, Griesbach, and Jung (2011), who show inconsistencies in the rating process during the European debt crisis.

methodology due to external influences, they apply the changed methodology in a similar way to all countries.

We proceed as follows: In the next two sections we briefly describe our data and methodology, before we summarize some of our main results in section 3 and conclude in section 4.

Data

In our study, we use panel data of European countries between 1995 and 2011 in order to analyze country ratings over time. In contrast to previous studies we use quarterly data to account for the short run dynamics in ratings that have been particularly striking in the recent financial turmoil. We employ two types of variables: firstly, risk measures such as ratings and bond returns, and secondly, risk explanatory variables such as government debt in percent of GDP, budget deficit, or inflation. Table 1 gives an overview of our explanatory variables.

Since our main dependent variable, the respective country rating, is published in letter categories typically ranging from AAA to D or C, we use a linear transformation to translate these ordered letter combinations into a scale from 21 (best rating) to 1 (worst rating). We collect rating histories from 1995 to 2011 for the three large agencies: Standard & Poor's, Moody's and Fitch Ratings. Since the different ratings follow the same trends, we use an average rating compiled from the three credit rating agencies for most parts of our analysis. Finally, we account for the short-term dynamics of the crisis by adjusting the average rating by $+0.3, 0$, or -0.3 respectively, depending on a positive, neutral, or negative outlook.

Methodology

We develop a simple macroeconomic model in order to explain variation in country ratings. Our main specification is a country fixed-effects regression of the following form:

$$R_{it} = \alpha_i + X'_{it}\beta + \varepsilon_{it} \tag{1}$$

Index i ($i = 1,\ldots, 15$) indicates the respective country while t (t=1995q1,. . . , 2011q4) stands for time. The dependent variable R_{it} is the linearly transformed and outlook-adjusted rating, while matrix X'_{it} contains macroeconomic explanatory variables. Unobserved heterogeneity in the model is absorbed by the country specific effect α_i. Lastly, ε_{it} is an error term with $\varepsilon_{it} \sim \mathcal{N}(0, \sigma^2)$. To account for heteroskedasticity and autocorrelation we use robust standard errors for the estimation of our model.

In order to examine whether ratings are consistent over time, we expand our model with time dummy variables and interaction terms. The dummy for the convergence-period from the introduction of the Euro until the outbreak of the financial crisis (1999q1−2007q4) is called "Euro", the dummy for the current crisis period (2008q1 − 2011q4) is termed "Crisis". Hence, we estimate the following extended model:

$$R_{it} = \alpha_i + \gamma Euro + \delta Crisis + X'_{it}\beta + (X_{it}xEuro)'\theta + (X_{it}xCrisis)'\eta + \varepsilon_{it} \tag{2}$$

Positive coefficients for interaction terms between macro variables and time dummies indicate a change in the rating methodology. Therefore, coefficients θ and η are of particular interest.

Results

Our main results are shown in Table 2. Column 1 starts with an estimation of our simple macroeconomic model specified in Equation (1). The influence of the distinct variables is not surprising: While higher debt levels, higher unemployment, higher inflation and a greater budget deficit correspond to worse ratings on average, a higher GDP per capita or a more efficient bureaucracy tend to increase the country rating. The adjusted R-squared shows that our model is able to explain over 90% of the variation in country ratings.

In a second step we include dummy variables for the period following the introduction of the Euro ($1999-2007$) and the current crisis period ($2008-2011$) in order to test whether rating agencies change their methodology over time due to external influences. Somewhat surprisingly, the dummy variable for the crisis period is positive and significant at the 1%-level. This challenges the common perception of too harsh downgrades between 2008 and 2011. To the contrary, a potential explanation for our results is that positive ratings were maintained for too long during the crisis, and that downgrades should have occurred much earlier (see Tichy (2011)). The insignificant coefficient for the convergence period between 1999 and 2008 indicates that the introduction of the common currency did not have a positive impact on the country ratings per se, but that it was accompanied by an improvement in fundamentals that justified the upgrades in the ratings of currency union members.

Results in column 2 indicate that country ratings are not always consistent over time. The same set of fundamentals might induce different ratings at different points in time. To test whether agencies also change the weight attributed to the distinct variables we include interaction terms between the dummy variables and the macroeconomic variables in column 3. Interaction terms for the convergence period are mostly insignificant and smaller than for the crisis period, which is why we concentrate on the latter and report differences between the pre- and the post-crisis period in column 4. Results hint at a shift in the weighting of macroeconomic and fiscal variables. For example, it seems as if the rating agencies put less emphasis on inflation rates and budget deficits and thus acknowledged the necessity of fiscal stimuli in the recent economic downturn. Conversely, the interaction term for the level of debt to gross domestic product has the same sign as the variable itself, indicating that this variable became more important during the crisis.

In the media as well as in academic publications (e.g. Gärtner, Griesbach, and Jung (2011)), rating agencies were harshly criticized for their treatment of Greece, Italy, Ireland, Portugal and Spain. It was suspected that the downgrades were arbitrary in the sense that they did not reflect the real economic situation but rather followed market sentiments. Our paper shows that this criticism is not valid. We examine the residuals produced by the fixed effects model with dummy and interaction terms and find that these are not systematically negative for the aforementioned countries in the crisis period. In other words, downgrades were justified by fundamental macroeconomic and fiscal data.

Conclusion

Summing up, our analysis suggests that rating agencies did not anticipate the deterioration of sovereign credit quality in the aftermath of the financial crisis and that their rating methodology was time inconsistent. However, we find no evidence that the GIIPS countries were subject to arbitrary downgrades. In the light of readily-available market assessments of default risk, such as CDS quotes, ratings appear increasingly redundant due to their delayed processing of viable information.

References

Afonso, A. (2003): “Unterstanding the Determinants of Government Debt Ratings: Evidence for the Two Leading Agencies,” *Journal of Economics and Fi-*

nance, 27(1), 53–74.

Afonso, A., P. Gomes, and P. Rother (2007): "What "Hides" behind Sovereign Debt Ratings?," *ECB Working Paper Series*.

Butler, A. W., and L. Fauver (2006): "Institutional Environment and Sovereign Credit Ratings," *Financial Management*, 35(3), 53–79.

Cantor, R., and F. Packer (1996): "Determinants and Impacts of Sovereign Credit Ratings," *Economic Policy Review*, 2(2), 37–53.

Deen, M. (2011): "ECB's Noyer Says Rating Companies Have Become "Irrational"," Accessed November 3, 2012.

Gärtner, M., B. Griesbach, and F. Jung (2011): "PIGS of Lambs? The European Sovereign Debt Crisis and the Role of Rating Agencies," *International Advances in Economic Research*, 17(3), 288–299.

Tichy, G. (2011): "Credit Ratings Agencies: Part of the Solution or Part of the Problem," *Intereconomics*, 46(5), 232–262.

Appendix

	Unit	Source	Frequency	Transformation
GDP per Capita	1000 € /Inhabitant	IMF	Anually	Cubic-Spline
Inflation	YoY in %	OECD	Quarterly	-
Unemployment	% annualized	OECD	Monthly	Average
Gross Public Debt	% of GDP	IMF	Anually	Linear Interpolation
Budget Deficit	% of GDP	IMF	Anually	Extrapolation
Government Effectiveness	-2.5 to + 2.5	World Bank	Semi-Annually	Linear Interpolation
10yr Government Bond Returns	%	ECB, Datastream	Monthly/ Daily	Average
CDS-Prices	Bps	Bloomberg, Datastream	Daily	Average

Table 1: *This table shows details for our set of variable (ratings are excluded here).*

	(1) Avg. Rating	(2) Avg. Rating	(3) Avg. Rating	(4) Avg. Rating
Gross Government Debt	-0.0142*** (0.003)	-0.0214*** (0.003)	-0.0207*** (0.004)	-0.0187*** (0.003)
GDP per Capita	0.0961*** (0.011)	0.0856*** (0.013)	0.0621*** (0.013)	0.0626*** (0.010)
Unemployment	-0.1851*** (0.017)	-0.1579*** (0.018)	-0.1694*** (0.017)	-0.1758*** (0.016)
Inflation	-0.3595*** (0.017)	-0.3502*** (0.017)	-0.4595*** (0.024)	-0.4130*** (0.018)
Primary Surplus	0.0407*** (0.010)	0.0578*** (0.010)	0.0734*** (0.028)	0.0357*** (0.013)
Government Effectiveness	0.3221*** (0.115)	0.5178*** (0.118)	0.4077*** (0.105)	0.4156*** (0.106)
Euro		-0.0390 (0.082)	-0.0441 (0.203)	
Crisis		0.4516*** (0.119)	0.6996*** (0.242)	0.6924*** (0.155)
Euro x Gross Public Debt			-0.0017 (0.002)	
Crisis x Gross Public Debt			-0.0176*** (0.003)	-0.0166*** (0.002)
Euro x Primary Surplus			-0.0418 (0.029)	
Crisis x Primary Surplus			-0.0957*** (0.031)	-0.0509*** (0.018)
Euro x Inflation			0.0858*** (0.028)	
Crisis x Inflation			0.3706*** (0.034)	0.3200*** (0.029)
Observations	831	831	831	831
Adjusted R-Squared	0.916	0.919	0.938	0.937
Country FE	YES	YES	YES	YES

Table 2: *The table shows different specifications of Equations (1) and (2). Independent variable is the country's average rating (over the three agencies) within a certain period, adjusted by the outlook. Standard are errors in parantheses.* *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.